Working Paper
Series

FINRISK
National Centre of Competence in Research
Financial Valuation and Risk Management

National Centre of Competence in Research
Financial Valuation and Risk Management

Working Paper No. 732

Is the Pricing Kernel U-Shaped?

Urs Schweri

First version: November 2011
Current version: November 2011

This research has been carried out within the NCCR FINRISK project on
"Behavioural and Evolutionary Finance"

Is the Pricing Kernel U-Shaped?*

Urs Schweri†

November 14, 2011

Abstract

There is strong empirical evidence that the pricing kernel is U-shaped, which provides a way to explain the substantial coskewness premium. Existing studies typically use a polynomial approximation of the pricing kernel. Problematically, these polynomials have, in most cases, increasing parts by construction. Therefore, it is not clear whether the increasing parts are an artifact of the chosen functional form. Taking this concept into consideration, this paper shows that pricing kernels, as estimated by the generalized method of moments on equity data, are still U-shaped and that the increasing part is not a statistical artifact. This conclusion derives from the fact that the functional form of kernels, which allows for strictly decreasing kernels as well as for kernels with increasing parts, is still U-shaped. These results arise from checking for higher order polynomials, various time horizons, and different functional forms of the kernel.

*The author would like to thank Wolfgang Härdle, Thorsten Hens, Alexandre Ziegler, Nilufer Caliskan, Rouslan Moro, and Christian Reichlin for valuable discussion. The financial support of the National Center of Competence in Research, 'Financial Valuation and Risk Management,' as managed by the Swiss National Science Foundation on behalf of the federal authorities is gratefully acknowledged.

†University of Zürich, Department of Banking and Finance, Plattenstrasse 32, 8032 Zurich, www.bf.uzh.ch, urs.schweri@bf.uzh.ch, +41 (0)44 634 48 22

1 Introduction

In the absence of arbitrage, a pricing kernel exists such that the price of an asset is equal to its pricing kernel-weighted payoffs. However, the no-arbitrage condition provides no information about the shape of the kernel (besides that of non-negativity). In equilibrium models with complete markets and a risk-averse representative agent who knows the probabilities of all states of the world, the pricing kernel is high in states with low resources because the marginal utility of one unit of additional consumption is high. However, in states with many resources, the pricing kernel is low. Therefore, the pricing kernel should decrease with resources.[1] In contradiction, there is, as shown in the following paragraphs, empirical evidence that the kernel is U-shaped within a certain range. However, studies such as those from Dittmar (2002), Potì (2006), and Post et al. (2008) do not investigate whether these increasing parts are significant. This paper tries to fill this gap.

Empirical evidence for increasing parts in the pricing kernel can be found in equity as well in option data. Estimating the kernel with equity returns yields a U-shaped pricing kernel, as Dittmar (2002), Potì (2006), and Post et al. (2008) have shown by approximating the kernel with a quadratic function or a higher order polynomial. The reason underlying this shape is the coskewness of single assets with the market portfolio returns (as a proxy for the available resources), such that the three-moment extension of the capital asset pricing model (CAPM) provides a significant risk premium for coskewness (Kraus and Litzenberger (1976), Friend and Westerfield (1980), Barone-Adesi (1985), Lim (1989), Harvey and Siddique (2000), Errunza and Sy (2005) and Smith (2007)). However, Dittmar (2002) and Post et al. (2008) show that the observed coskewness premium can no longer be explained if nonsatiation, risk aversion, and nonincreasing absolute risk aversion are imposed on the utility function of the representative agent (thereby translating into restrictions on the pricing kernel). Furthermore, Potì and Wang (2010) showed that unconstrained quadratic or higher order kernels imply relative risk aversions of above five for the representative investor, which is generally considered to be implausible. The pricing kernel in the well-known CAPM, discussed by Sharpe (1964), Lintner (1965), and Mossin (1966), is linear in relation to market returns, whereas the three-moment CAPM extends this with a quadratic term. On extending the CAPM by further moments, Fang and Lai (1997) and Hung (2008) found that co-kurtosis is also a relevant pricing factor. Considering a third order polynomial for the pricing, it still remains U-shaped in market returns, as shown by Dittmar (2002). Estimations of the

[1] Suitable textbook references are Magill and Quinzii (1996) or Cochrane (2001).

kernel from equity data clearly point to a U-shaped kernel.

Pricing kernel estimations done with option data also show a U-shaped kernel around the current stock price. Prominent examples of this are Aït-Sahalia and Lo (2000), Jackwerth (2000), and Rosenberg and Engle (2002). Note that estimating pricing kernels with option data allows us to estimate the pricing kernel within a broader range of market portfolio returns than is the case for equity data. The reason for this is that options exist with quite extreme strike prices, whereas such extreme values can rarely be observed in markets. Therefore, it turns out that pricing kernels typically fall after an increasing part of the kernel. If the view is restricted to the range of the kernel, which can be estimated by equity data, the kernel estimated by option data has a U-shape. More recent evidence is less clear: Detlefsen et al. (2007) estimated the kernel for German data at several points in time; in some cases, he derived a U-shape around the actual index level, and sometimes the kernel was merely decreasing. In another work, Golubev et al. (2008) provide some evidence that the increasing parts of kernels estimated using option prices are statistically significant. However, by using an asymmetric GJR-GARCH model with empirical innovations for option data, Barone-Adesi et al. (2008) and Barone-Adesi and Dall'O (2010) showed that the increasing parts of the kernel largely disappear.

In conclusion, there is empirical evidence for U-shaped kernels from equity as well as from option data around the current market price. Given that the estimations with equity data typically employ polynomial kernels and particularly quadratic kernels, the pricing kernel, by construction, has to be U-shaped (an inverse U-shape is also conceivable). This leads to the question as to whether the increasing part of the pricing kernel is merely a consequence of the polynomial functional form. Another interesting question is whether many observations of market returns are present in the increasing part. If not, then the increasing part could be a meaningless artifact. Unfortunately, the existing literature is not particularly helpful in addressing these concerns. For instance, Dittmar (2002), Potì (2006) and Post et al. (2008) estimated kernels with increasing parts and found that the coefficients before the polynomial terms were significant. However, their studies did not locate the minimum of the kernel or the origin of the increasing part of the kernel. Further, they do not provide any information on how often the economy lies in the increasing part of the kernel. Other studies, such as Hansen and Singleton (1982, 1983), and many subsequent works, have chosen a functional form (typically a power utility) such that the pricing kernel declines with any parameterization. More recently, Post and van Vliet (2006) and De Giorgi and Post (2008) estimated pricing kernels based on second-order stochastic dominance—and by that, assuming decreasing kernels—and found that the pricing kernels are steep in losses and flat in gains. However, this approach is also unhelpful in answering the question whether the pricing kernel is U-shaped. However, restricting kernels to be decreasing would result in such flat parts of the kernel.

The purpose of this paper is, therefore, to check whether the pricing kernel has increasing parts and, if so, where these are. The main contributions of this paper are as follows: First, it makes clear that the increasing parts of the kernel are not an artifact of the polynomial functional forms by estimating the kernel for functional forms, where the kernel may be U-shaped or where it is not (i.e., the piecewise linear kernel and the modified quadratic kernel, which starts at one point to be linear). These various functional forms, furthermore, allow direct testing for the increasing parts in the kernels. Second, estimating a higher order polynomial kernel than the literature reveals also that the dataset, including the Fama-French value, size, and momentum portfolios, has a clearly U-shaped kernel, which was not the case with a quadratic kernel.

Following a large part of the previously cited literature, the kernel is estimated on equity data by the generalized method of moments (GMM). Following Post et al. (2008), the pricing kernel is assumed to be constant over time. There are two main reasons for this: First, the question to be answered in this paper is if the pricing kernel is persistently U-shaped. Therefore, even if the pricing kernel is time varying, the focus is on the shape of the kernel as an average over the long run. In line with that is the usage of equity data, because the available data history is much longer for them than for option data. Second, theory does not show how the time variation of the kernel should be modeled. Taking, for example, the equilibrium model provided later in this paper, the pricing kernel depends only on the preferences of the representative agent. Given that the preferences remain roughly constant over time, the pricing kernel is also. Furthermore, it is reasonable to choose a simple econometric model, which is capable of modeling the main features but is only slightly misspecified, rather than a complicated one, where one cannot be certain that it is even more misspecified. Concerning the results of the estimations, the quadratic kernel turns out be U-shaped, in line with Dittmar (2002), Potì (2006) and Post et al. (2008). Each of them estimates the kernel on one dataset. To ensure robustness of the results, this paper estimates everything using five different datasets. Additionally, this paper quantifies the effect of the increasing parts of the kernel further: For two of these datasets, the kernel lies in the increasing region for more than one-quarter of the total observed period. In the other three datasets, the kernel lies in the increasing part for less than 2.5% of the observed periods. These results demonstrate two things: the pricing kernel is U-shaped, but the number of observations, when an increasing kernel occurs, may vary substantially.

An obvious issue with a polynomial kernel, especially a quadratic kernel, is that it almost automatically has increasing parts in it. Therefore, it is interesting to see how well a kernel does that is restricted to be decreasing. For example, Post et al. (2008) restricted the parameters of the kernel in such a way that on all observed data, the kernel is falling. With such an approach, the parameters may be strongly restricted: in my dataset the observed market returns range from -0.29 to 0.384, but 95% of the market returns are observed between -0.105 and 0.1. If a falling kernel is enforced in too large a range, the parameters of the whole kernel in that setup are massively more restricted. This becomes problematic if, as shown, only 5% of the observations can be found on approximately one-half of the restricted range.

A further step for investigating the shape of the kernel is to restrict the slope of the kernel to zero in increasing areas (as in Dittmar (2002)). Since the kernels with the restricted and the unrestricted slope are not in this way nested in each other, it is noteworthy that the nonincreasing kernel fits the data worse than the kernel with the increasing parts. However, this effect is less strong than in Dittmar (2002). Setting the slope of the increasing parts to zero is one possible way, but might it be better to restrict the slope to be smaller than another level, for example -0.01? An innovation of this paper is to estimate the level where the slope of the kernel should be restricted: it turns out that this level, in four out of five datasets, is (insignificantly) positive. That is, even if the functional form of the kernel explicitly allows for the increasing parts of the kernel to be flat, the estimation shows increasing parts.

A potential issue is the order of the polynomial; the literature, for example Dittmar (2002) and Potì (2006), stops with pricing kernels of order 3. This paper tests polynomials up to order 7. In four out of our five datasets, this is enough. However, in the fifth dataset, the Fama-French value, size, and momentum portfolios, a pricing kernel with a polynomial of at least order 6 with a clear U-shape is appropriate. Chung et al. (2006) and Nguyen and Puri (2009) showed that Fama-French value, size, and momentum excess returns can be explained by a higher-order polynomial of the market excess return; the resulting pricing kernel is, therefore, a consequence of this. Because a polynomial of order 3 is sufficient, if the momentum portfolios are not included, it can be concluded that this effect can be attributed to the momentum portfolios.

Finally, a polynomial may not be the correct functional form of the pricing kernel. To check this, a piecewise linear kernel is estimated. It turns out that the estimated piecewise linear pricing kernel has approximately the same shape as the polynomial kernel, including its increasing parts.

Overall, this paper shows, with an extremely broad range of different tests, that the pricing kernels have a U-shape and that this shape is neither the result of a misspecified functional form nor a statistical artifact. Nonetheless, one word of caution: the estimation of the kernel cannot be done very precisely. The observed substantial variation between the kernels estimated from the different datasets and the poor significance of most statistical tests demonstrate this. Nonetheless, estimations on five different datasets and over a time horizon of more than 80 years point to a U-shape for the kernel. Hence one can be confident that the increasing parts of the kernel exist and that they, therefore, should be taken into account for asset pricing.

The remainder of the paper is arranged as follows. Section 2 provides the model framework and section 3 the estimation methodology. Section 4 describes the datasets, and the pricing kernels are estimated and tested in section 5. The effects of these kernels on the utility function of a rational, representative investor are shown in section 6. Finally, section 7 presents the conclusions.

2 Model framework

No arbitrage implies the existence of some risk-neutral measure π, such that for all assets k, the expected return under π is the risk-free rate R^f (Harrison and Kreps (1979))—that is,

$$R^f = \mathbb{E}_\pi(R^k) \qquad\qquad \text{for all assets } k.$$

Let p_s denote the physical probability of state $s \in \{1, \ldots, S\}$. Then,

$$R^f = \mathbb{E}_\pi(R^k) = \sum_{s=1}^{S} p_s \frac{\pi_s}{p_s} R_s^k = \sum_{s=1}^{S} p_s L_s R_s^k = \mathbb{E}_P(LR^k), \tag{1}$$

where $L_s = \frac{\pi_s}{p_s}$ is the pricing kernel.[2] To keep notation simple, the physical probability $\mathbb{E}_P$ is, in the future, written as $\mathbb{E}$. As Equation (1) holds for any

[2]Often—for example, in Cochrane (2001)—the stochastic discount factor is used instead of the pricing kernel. Both measures for the state prices differ in a (multiplicative) constant.

assets k and j, one has

$$0 = \mathbb{E}\left(L \cdot (R^k - R^j)\right), \tag{2}$$

and furthermore, because π and p are probabilities and $L \geq 0$,

$$\mathbb{E}(L) = 1. \tag{3}$$

No arbitrage implies these conditions for the pricing kernel. Instead of an unconditional expected value, a conditional expected value—that is, $0 = \mathbb{E}\left(L(R^k - R^j)|\Omega_t\right)$, as in Dittmar (2002) and Potì (2006), for example—can be used, where Ω_t is the information available in period t. More precisely, Ω_t are the realizations of some random variables in t, on which the conditional expected value is defined. How the conditioning variables included in Ω_t should be chosen is unclear. Cochrane (2001, p. 145) summarized the issue as follows: "The situation is not repaired by simple inclusion of some conditioning information. Models such as the CAPM imply a conditional linear factor model with respect to investors' information sets. However, the best we can hope to do is to test implications conditioned down on variables that we can observe and include in a test. Thus, a conditional linear factor model is not testable." A possible way to circumvent this is to take the expected value of the conditional expected value

$$0 = \mathbb{E}\left(\mathbb{E}\left(L \cdot (R^k - R^j)|\Omega_t\right)\right) = \mathbb{E}\left(L \cdot (R^k - R^j)\right).$$

That is, the conditional model implies the unconditional one if the pricing kernel, L, is constant over time. Thus, it can be concluded that the unconditional model must hold, in any case, in the long run. While this method is probably not the most efficient way to estimate a pricing kernel, it relies on only a few assumptions, and the results cannot be influenced by wrongly chosen conditional variables.

2.1 Relation to the utility function of a representative investor

No arbitrage implies the existence of a positive pricing kernel. To obtain more information on the shape of the pricing kernel and on which variables the kernel depends, an equilibrium model can be used. Consider a two-period model in which a representative agent has initial wealth w_0 and an increasing, strictly concave, and differentiable utility function. Then the representative agent maximizes his utility:

$$\max_{\lambda_k} u(c_0) + \delta\mathbb{E}(u(C)),$$

where δ is the time discount. The maximization is done under the following budget constraints:

$$C = w_0 \left(\sum_{k=1}^{K} \lambda_k R^k + \left(1 - \sum_{k=1}^{K} \lambda_k \right) R_0 \right)$$
$$c_0 = w_0 \left(1 - \sum_{k=1}^{K} \lambda_k \right),$$

where c_0 is consumption in the first period, as given by the initial wealth minus the investment, into a portfolio of assets where λ_k is the portfolio weight of asset $k \in \{1, \ldots, K\}$, and C_s is consumption in state s, given by the payoffs of the portfolio bought in the first period. Then the budget constraints are inserted into the maximization problem. The first-order conditions of the utility maximization problem then imply for all assets k

$$R^f = \mathbb{E} \left(\delta \frac{u'(C)}{u'(c_0)} R^k \right).$$

A comparison with equation (1) immediately shows that the pricing kernel is given by

$$L = \delta \frac{u'(C)}{u'(c_0)}. \tag{4}$$

In the previous model, δ and c_0 are constants. The likelihood ratio process is, therefore, proportional to the marginal utility in the next period. Further, it is falling in consumption because of the concave utility function; that is, the increasing parts in the likelihood ratio process cannot be explained by this model.

The estimated pricing kernel supports the model as long as the pricing kernel is nonincreasing. However, empirical evidence shows that there may be increasing parts in the pricing kernel (see, for example, Aït-Sahalia and Lo (2000) for evidence from option data or Dittmar (2002) for evidence based on equity data). In this situation, any assumption of the model may be violated. For instance, Ziegler (2007) identifies problems of aggregation, misestimated beliefs of the agents, Peso problems, and heterogeneous beliefs as possible reasons for the observed increasing parts in the pricing kernel. However, several examples and some empirical evidence show that none of these explanations may suffice to explain the observed increasing parts of the kernel under a reasonable set of assumptions. Hens and Reichlin (2010) show, furthermore, with simple examples that a nonconcave utility function of the representative investor, incomplete markets, or heterogeneous beliefs may explain the increasing parts in the kernel. In this setup, the latest solution seems to be the best explanation since the other two possibilities need unrealistic assumptions or the results are fragile if the parameters are changed a bit. Over all, it is challenging to explain the increasing parts of the kernel theoretically under plausible assumptions.

With the equilibrium argument above, the pricing kernel depends on the consumption of the representative investor. This consumption is typically approximated by the use of either aggregate consumption or market portfolio return data. The latter source assumes that the only source of income for an investor is his assets and that by market clearing, the representative investor holds all assets in equilibrium and, therefore, earns the market return. Mankiw and Shapiro (1986) have shown that using market portfolio returns as a consumption proxy explains the observed risk premia (in a CAPM setup) much better than does using aggregated consumption. The main competitor for market portfolio returns—namely, aggregate consumption data from the National Income and Products Accounts, which has been used by Hansen and Singleton (1982) and many others—causes several problems. For example, Breeden et al. (1989), Ferson and Harvey (1992), Wilcox (1992), and Slesnick (1998) discuss measurement errors, definitional problems, issues with seasonal adjustment, and other problems with the aggregation of the consumption data over time. Furthermore, many behavioral explanations—for example, narrow framing, loss aversion, or mental accounting—demonstrate why wealth (or changes in it) should be included in the utility function of the representative agent and also, therefore, in the pricing kernel. St-Amour (2007) gives a short overview of this literature. For example, Barberis and Huang (2006) and Hens and Wöhrmann (2006) demonstrated that the equity premium puzzle can be explained in that way. For these reasons, the pricing kernel is chosen as a function of the market excess return: $L(r_m)$ with $r_m = R^M - R^f$. A side effect of this is that by subtracting the riskless rate, the kernel is in real terms. This makes sense because the consumer is not interested in his nominal wealth but in the amount of real consumption he can afford with his wealth. Nonetheless, the impact of stochastic inflation on portfolio decisions may be complex, as demonstrated by Brennan and Xia (2002) in a continuous time setup. The chosen approach tries, therefore, to keep the impact of stochastic inflation on our results as small as possible but does not claim to solve this issue.

2.2 Functional form of the pricing kernel

To test for increasing parts, kernels that can, but do not have to, contain increasing parts are especially interesting. In the following, we consider linear, polynomial, and piecewise linear kernels. The first two types are important since they are used in a large part of the literature. However, many polynomial, and especially quadratic, kernels have the disadvantage that almost by definition they have increasing parts. Because of that, the piecewise linear kernels are also used.

In the CAPM, the pricing kernel is a linear function of market excess return. Rubinstein (1973) showed that the CAPM can be considered as a first-order Taylor approximation of a representative investor with an arbitrary utility function. A higher-order Taylor approximation of the pricing kernel $L = \delta u'(C)/u'(c_0) = \delta u'(r_m)/u'(c_0)$ at the risk-free rate (assuming that c_0 is constant) has the following form:

$$L(r_m) = h_0 + h_1 u'' \cdot r_m + h_2 u''' \cdot r_m^2 + h_3 u'''' \cdot r_m^3 + \ldots$$

The quadratic and cubic terms can be interpreted as preferences for skewness and kurtosis. Typically, most people prefer positively skewed distributions without fat tails; therefore, $h_2 u'''$ should be positive, and $h_3 u''''$ should be negative. The polynomial kernel is defined as

$$L(r_m) = \theta_0 + \theta_1 \cdot r_m + \theta_2 \cdot r_m^2 + \theta_3 \cdot r_m^3 + \ldots = \theta_0 + \sum_j \theta_i \cdot r_m^j. \qquad (5)$$

This kernel has two advantages: it is extremely general (every continuous function can be approximated by it) and it can be written in terms of linear factors, where $r_m, r_m^2, \ldots$ are the factors. However, is a polynomial the right type of function to approximate a pricing kernel? By construction, a Taylor approximation describes a function well at the point of approximation, but worsens the further it is away from that point. To estimate the pricing kernel for large or small market returns, other functional forms could potentially work better. An alternative is to use a piecewise linear kernel with breakpoints $q_1, \ldots, q_n$, i.e., as follows:

$$L(r_m) = \tau_0 + \tau_1 r_m + \begin{cases} 0 & \text{for } r_m < q_1 \\ \tau_2(r_m - q_1) & \text{for } q_1 \leq r_m < q_2 \\ \tau_2(r_m - q_1) + \tau_3(r_m - q_2) & \text{for } q_2 \leq r_m < q_3 \\ \vdots & \vdots \\ \tau_2(r_m - q_1) + \tau_3(r_m - q_2) + & \\ \quad \ldots + \tau_{n+1}(r_m - q_n) & \text{for } q_n \leq r_m. \end{cases} \qquad (6)$$

The main advantage of this functional form is the enormous number of possible shapes of the kernel. Post and van Vliet (2006) is one of the few studies where a piecewise linear marginal utility function (i.e., in a representative agent model, a piecewise linear pricing kernel) has been used. The authors mainly find that the market portfolio is not mean variance efficient but that third-order stochastic dominance seems to hold for the market portfolio. Since they are focusing on a linear program to check the stochastic dominance, they focus on decreasing kernels. The main reason for the rare usage of the piecewise linear kernel may be that the first derivative in r_m is not continuous in all points. For the generalized method of moments (GMM) estimation, it is required only that the piecewise linear pricing kernel is differentiable in all τ, which is obviously given. Nevertheless, for numerical optimization the noncontinuous first derivative of r_m may be problematic. The estimation methods are discussed in more detail in the next section.

3 Estimation methods

In a further step, the pricing kernel must be estimated from data. This can be done in various ways: Based on no-arbitrage, this step is especially easy for a linear kernel, which can be determined out of the market portfolio and the risk-free asset, and it is then identical with the famous CAPM. If it is possible to represent the kernel in a linear form in factors, then the kernel can be estimated via OLS regressions. More complicated kernels may be estimated by GMM. All these estimation methods are discussed in more detail in the following sections.

3.1 Benchmark CAPM

The CAPM of Sharpe (1964), Lintner (1965) and Mossin (1966) assumes a linear pricing kernel $L = \tilde{\theta}_0 + \tilde{\theta}_1 r_m$ with market excess return $r_m = R^M - R^f$. This pricing kernel can easily be estimated out of the return of the market portfolio and the riskless asset. Assuming that asset k in equation (2) is the market portfolio and that asset j is the risk-free rate, the following conditions are satisfied:

$$
\begin{aligned}
0 &= \mathbb{E}\left(L \cdot (R^M - R^f)\right) = \tilde{\theta}_0 \mathbb{E}(r_m) + \tilde{\theta}_1 \mathbb{E}(r_m^2) \\
1 &= \mathbb{E}(L) = \tilde{\theta}_0 + \tilde{\theta}_1 \mathbb{E}(r_m).
\end{aligned}
$$

Solving this system of two linear equations for $\tilde{\theta}_0$ and $\tilde{\theta}_1$ results in the following:

$$\tilde{\theta}_0 = \frac{\mathbb{E}(r_m^2)}{\operatorname{var}(r_m)} \qquad \text{and} \qquad \tilde{\theta}_1 = -\frac{\mathbb{E}(r_m)}{\operatorname{var}(r_m)}. \tag{7}$$

Plugging in the average and variance of past market portfolio excess returns is the simplest way to estimate the parameters of a linear pricing kernel. The advantage of this method of estimating a pricing kernel is that it depends only on the risk-free rate and the return of the market portfolio. Therefore, it does not depend on the returns of individual assets. This is an advantage, because it is typically not feasible to include every single asset in the world in an empirical study; therefore, the choice of the assets to include may affect the results. For the rest of the paper, this estimation method for a pricing kernel is referred to as the benchmark CAPM model. The following two subsections describe two additional methods for estimating more general pricing kernels.

3.2 Factor models

If pricing kernels are linear combinations of factors, they can be estimated using linear factor models. The CAPM (or its higher moment versions) or the Fama-French three-factor model are special cases of factor models. Furthermore, all the functional forms of the previous section can be rewritten in factor form. Since a large part of the literature focuses on factor models and on the risk premia for the different factors, it is helpful to illustrate their link with the pricing kernel. Assume that the pricing kernel is given by $L = b_0 + \boldsymbol{b}'\boldsymbol{f}$, where $\boldsymbol{f}$ is a vector of factors that vary over time, and b_0 and the vector b are constants. Given $\mathbb{E}(L) = 1$, this can be rewritten as $L = 1 + \boldsymbol{b}'(\boldsymbol{f} - \mathbb{E}(\boldsymbol{f}))$. Writing everything in terms of excess returns, $R_k^e = R^k - R^f$, Equation (2) becomes

$$0 = \mathbb{E}\left(L \cdot (R^k - R^f)\right) =: \mathbb{E}\left(L \cdot R_k^e\right).$$

With $\boldsymbol{R^e}$ defined as the vector of excess returns of all assets, this is equivalent to

$$\begin{aligned} 0 &= \mathbb{E}\left(L \cdot \boldsymbol{R^e}\right) = \mathbb{E}(\boldsymbol{R^e}) + \boldsymbol{b}' \operatorname{cov}\left(\boldsymbol{f}, \boldsymbol{R^e}\right) \\ \mathbb{E}(\boldsymbol{R^e}) &= -\boldsymbol{b}' \operatorname{cov}\left(\boldsymbol{f}, \boldsymbol{R^e}\right) = -\boldsymbol{b}' \operatorname{var}(\boldsymbol{f}) \operatorname{var}(\boldsymbol{f})^{-1} \operatorname{cov}(\boldsymbol{f}, \boldsymbol{R^e}) \\ &= \boldsymbol{\lambda}'\boldsymbol{\beta}, \end{aligned}$$

where

$$\boldsymbol{\lambda} = -\operatorname{var}(\boldsymbol{f})\boldsymbol{b} \qquad \text{and} \qquad \boldsymbol{\beta} = \operatorname{var}(\boldsymbol{f})^{-1}\operatorname{cov}(\boldsymbol{f}, \boldsymbol{R^e})$$

are the risk premium and risk exposure of every risk factor. The CAPM is a special case of that model. If the CAPM pricing kernel $L = \tilde{\theta}_0 + \tilde{\theta}_1 r_m$ is plugged in, the expected excess return of an asset in the CAPM is a risk premium, $\tilde{\lambda} = \mathbb{E}(r_m)$,[3] times the CAPM-$\beta$. It measures the exposure to the market risk and is defined by $\tilde{\beta} = \operatorname{cov}(r_m, R^e)/\operatorname{var}(r_m)$.

In general, $\boldsymbol{\beta}$ are the estimated multiple regression coefficients of excess returns R^e on the demeaned factors $\boldsymbol{f}$. This offers one way to estimate $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$: first, regress $\boldsymbol{R^e}$ on the demeaned factors $\boldsymbol{f}$ to obtain the estimator $\hat{\beta}$. Second, regress the average excess returns of the assets $\bar{\boldsymbol{R}}^{\boldsymbol{e}}$ on $\hat{\boldsymbol{\beta}}$ to obtain an estimator for the risk premium $\hat{\boldsymbol{\lambda}}$. Most of the extant literature stops at this point and estimates the risk premia, λ_k, of each factor. Because $\boldsymbol{b} = -\operatorname{var}(\boldsymbol{f})^{-1}\boldsymbol{\lambda}$, the estimator for $\boldsymbol{b}$ is

$$\hat{\boldsymbol{b}} = -\left(\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{f_t} - \frac{1}{T}\sum_{t=1}^{T} f_t\right)\left(\boldsymbol{f_t} - \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}\right)'\right)^{-1}\hat{\boldsymbol{\lambda}}.$$

The estimated pricing kernel, $\hat{L}$, is obtained by plugging $\hat{\boldsymbol{b}}$ into the definition of the pricing kernel:

$$\hat{L} = 1 + \hat{\boldsymbol{b}}'\left(\boldsymbol{f_t} - \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}\right).$$

Appendix A shows that under weak assumptions (heteroskedasticity is allowed), $\hat{\boldsymbol{b}}$ is a consistent estimator of $\boldsymbol{b}$. Through the consistency of $\hat{\boldsymbol{b}}$ and the law of large numbers, it directly follows that $\hat{L}$ is consistent. While consistency alone does not provide any information about confidence intervals or the significance of the parameters, these can be obtained using GMM estimation.

3.3 Estimation via the generalized method of moments

Compared to the factor model, GMM permits more general pricing kernels and provides asymptotic test statistics for the estimated parameters. Assuming that the pricing kernel is a function of the market excess return r_m,

[3] To obtain that, plug equation (7) into the definition of λ.

equations (2) and (3) imply the following moment restrictions:

$$0 = \mathbb{E}\left(L(r_m) \cdot (R^k - R^f)\right)$$
$$0 = \mathbb{E}\left(L(r_m)\right) - 1.$$

The market portfolio also has to be priced correctly. Therefore,

$$0 = \mathbb{E}\left(L(r_m) \cdot r_m\right) = \sum_k \mathbb{E}\left(L(r_m) \cdot \omega_k (R^k - R^f)\right),$$

where ω_k is the weight of asset k in the market portfolio in every time period. Moreover, this condition ensures that the sum of the pricing error over all assets is close to zero. Given these moment conditions, the parameter of the pricing kernel can be estimated using GMM. Note that no assumptions about the distribution of the returns are required, only that all moments must exist. Additionally, some regularity conditions should be satisfied.

Assume that $\boldsymbol{X}_t$ is a vector of the data needed to estimate the model (mainly asset and market returns in t); then $\boldsymbol{\theta_T}$ is the vector of the true parameter of the pricing kernel, and the vector of all moment conditions is written as

$$0 = \mathbb{E}(\boldsymbol{g}(\boldsymbol{X}_t, \boldsymbol{\theta_T})) := \mathbb{E}\begin{pmatrix} L(r_m, \boldsymbol{\theta_T}) \cdot (R^1 - R^f) \\ \vdots \\ L(r_m, \boldsymbol{\theta_T}) \cdot (R^K - R^f) \\ \sum_k L(r_m, \boldsymbol{\theta_T}) \cdot \omega_k (R^k - R^f) \\ L(r_m, \boldsymbol{\theta_T}) - 1 \end{pmatrix}.$$

The estimated parameter $\hat{\boldsymbol{\theta}}$ is chosen such that the deviations from the moment conditions are minimized (the deviations are weighted by a weighting matrix $\boldsymbol{W}$):

$$\hat{\boldsymbol{\theta}}(\boldsymbol{W}) = \operatorname{argmin}_{\boldsymbol{\theta}} n \boldsymbol{g_n}(\boldsymbol{\theta})' \boldsymbol{W} \boldsymbol{g_n}(\boldsymbol{\theta})$$
$$\boldsymbol{g_n}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{g}(\boldsymbol{X}_t, \boldsymbol{\theta}).$$

Under some regularity conditions, the estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{W})$ is consistent, and if $\boldsymbol{W}$ is a consistent estimator of the inverse covariance matrix of $\boldsymbol{g}(\boldsymbol{X}_t, \boldsymbol{\theta})$, the estimator is asymptotically efficient. The difference between the estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{W})$ and the true parameter $\boldsymbol{\theta_T}$ is asymptotically normally distributed. This approach allows all kinds of statistical tests, especially t-tests.

The minimization problem of the GMM estimator is solved numerically. Therefore, accurate starting values are crucial. Because the factor model presented before provides a consistent estimator of the model parameters, I use them as starting values.

To obtain an estimation of $\boldsymbol{W}$, the two-step GMM method in Hansen (1982) is applied.[4] Using this method, the model is first estimated with $\boldsymbol{W}$ as the identity matrix and the parameter estimates from the factor regression as starting values. Then a heteroskedasticity and autocorrelation consistent (HAC) covariance matrix for $\boldsymbol{g}(\boldsymbol{X}_t, \boldsymbol{\theta})$ is calculated (see Newey and West (1987)). With the inverse of this covariance matrix serving as the weighting matrix $\boldsymbol{W}$, the definitive model is estimated. Hansen et al. (1996) suggested more sophisticated GMM estimators: the iterated GMM and continuously updated GMM estimator, which differ in the way they determine $\boldsymbol{W}$. Newey and Smith (2004) and Anatolyev (2005) concluded that the two-step and iterated estimators are asymptotically equivalent and that the continuously updated estimator has a smaller asymptotic second-order bias than the other two estimators. With finite samples, these results can obviously differ. In this paper, the two-step estimator is used because it turned out to be the numerically most robust estimator. The iterated and continuously updated estimator leads to extremely volatile pricing kernels. (Chapman (1997) found similar problems with the iterated GMM estimator, in which the observed average returns are far from the mean returns predicted by the model.)

4 Data

The suggested estimation methods need the excess returns (i.e., $R^k - R^f$) of various assets and the market, and the market capitalizations of all the assets as well. Monthly data are used, starting in 1926, from the Center for Research in Security Prices (CRSP) at the University of Chicago. The CRSP all-share index, a value-weighted index of all common stocks listed on the NYSE, AMEX, and NASDAQ markets, is taken as a proxy for the market portfolio. The risk-free rate is the one-month T-Bill rate. To avoid spurious results, all kernels are estimated using the monthly excess returns of five different sets of data. The first two sets of assets are the 17 and 30 Fama-French industry portfolios.[5] They group all NYSE, AMEX, and NASDAQ stocks

[4] All estimations are done with R. For the GMM estimations, the GMM-package of Chaussé (2010) has been used.

[5] Fama and French used data from CRSP to calculate their returns. Compustat data are used for the portfolio weights of the value portfolios.

				17 Industry Portfolios						
Industry	N	Mean	SD	Skew.	Kurt.	β	γ	δ	ACF(1)	p
Food	1000	0.0069	0.049	-0.02	9.4	0.8	2.2	0.8	0.09	0.005
Mines	1000	0.0071	0.068	-0.12	5.3	0.9	1.1	0.7	0.07	0.028
Oil	1000	0.0078	0.061	0.29	7.0	0.9	3.1	0.9	0.01	0.843
Clths	1000	0.0057	0.062	0.33	8.2	0.9	2.5	0.8	0.17	0.000
Durbl	1000	0.0060	0.078	1.36	19.7	1.3	7.3	1.5	0.21	0.000
Chems	1000	0.0074	0.064	0.34	9.5	1.0	4.2	1.1	0.11	0.001
Cnsum	1000	0.0072	0.050	0.25	9.2	0.7	2.8	0.8	0.06	0.070
Cnstr	1000	0.0064	0.069	0.43	8.8	1.2	4.7	1.2	0.11	0.001
Steel	1000	0.0065	0.086	1.37	16.7	1.4	9.1	1.6	0.11	0.000
FabPr	1000	0.0060	0.061	0.15	9.1	1.0	3.6	1.0	0.10	0.001
Machn	1000	0.0078	0.072	0.18	8.6	1.2	4.6	1.2	0.11	0.000
Cars	1000	0.0078	0.079	1.15	16.7	1.2	6.7	1.5	0.15	0.000
Trans	1000	0.0063	0.072	1.01	15.2	1.2	6.1	1.3	0.15	0.000
Utils	1000	0.0056	0.057	0.13	10.5	0.8	2.8	0.9	0.11	0.000
Rtail	1000	0.0068	0.060	-0.01	8.1	0.9	2.7	1.0	0.14	0.000
Finan	1000	0.0071	0.070	0.56	14.3	1.2	5.1	1.3	0.17	0.000
Other	1000	0.0056	0.052	-0.19	6.8	0.9	1.8	0.8	0.12	0.000
Market	1000	0.0091	0.055	0.14	10.5	1.0	3.5	1.0	0.12	0.000

Table 1: Summary statistics of the monthly excess returns (i.e. $R^k - R^f$) of the 17 industry portfolios and the market portfolio: average; standard deviations; skewness; kurtosis; CAPM-β; γ, the exposure to the skewness risk (see equation 8); δ, the exposure to the kurtosis risk (see equation 9) and the first-order autocorrelation coefficient (with p-value) of portfolio excess returns.

Industry	N	Mean	SD	Skew.	Kurt.	β	γ	δ	ACF(1)	p
				30 Industry Portfolios						
Food	1000	0.0068	0.049	0.05	9.4	0.7	2.3	0.8	0.08	0.007
Beer	1000	0.0093	0.075	1.84	25.2	1.0	3.0	1.0	0.09	0.005
Smoke	1000	0.0084	0.059	0.07	6.4	0.6	2.4	0.7	0.07	0.035
Games	1000	0.0076	0.091	0.64	12.3	1.4	6.3	1.6	0.19	0.000
Books	1000	0.0060	0.071	0.52	9.7	1.1	5.0	1.2	0.18	0.000
Hshld	1000	0.0063	0.061	0.37	15.5	0.9	3.7	1.1	0.08	0.008
Clths	1000	0.0055	0.061	0.30	7.9	0.8	1.9	0.6	0.15	0.000
Hlth	1000	0.0077	0.058	0.18	10.1	0.9	3.4	0.9	0.08	0.018
Chems	1000	0.0073	0.064	0.37	9.7	1.0	4.3	1.1	0.10	0.001
Txtls	1000	0.0063	0.081	1.05	12.6	1.2	6.3	1.3	0.18	0.000
Cnstr	1000	0.0061	0.070	0.36	8.9	1.2	4.3	1.1	0.13	0.000
Steel	1000	0.0065	0.085	1.37	16.7	1.4	9.1	1.6	0.11	0.000
FabPr	1000	0.0073	0.073	0.48	10.4	1.2	5.4	1.3	0.13	0.000
ElcEq	1000	0.0088	0.078	0.60	11.6	1.3	6.1	1.4	0.10	0.001
Autos	1000	0.0076	0.081	1.23	17.4	1.2	6.9	1.5	0.15	0.000
Carry	1000	0.0080	0.078	0.49	8.4	1.2	5.0	1.2	0.11	0.001
Mines	1000	0.0067	0.073	0.13	6.7	0.9	3.1	1.0	0.05	0.108
Coal	1000	0.0098	0.092	0.87	9.8	0.8	0.1	0.3	0.04	0.217
Oil	1000	0.0077	0.061	0.29	7.0	0.9	3.1	0.9	0.01	0.829
Util	1000	0.0056	0.057	0.13	10.5	0.8	2.8	0.9	0.11	0.000
Telcm	1000	0.0051	0.046	0.00	6.2	0.7	1.2	0.6	0.09	0.007
Servs	1000	0.0089	0.086	1.11	19.2	0.8	-1.6	0.4	0.03	0.361
BusEq	1000	0.0080	0.069	-0.22	6.1	1.1	2.1	0.9	0.09	0.005
Paper	1000	0.0071	0.061	0.36	9.5	1.0	3.9	1.0	0.06	0.071
Trans	1000	0.0059	0.073	1.10	16.0	1.1	6.3	1.3	0.15	0.000
Whlsl	1000	0.0052	0.075	0.65	14.3	1.1	4.2	1.2	0.19	0.000
Rtail	1000	0.0069	0.060	0.02	8.0	0.9	2.8	1.0	0.14	0.000
Meals	1000	0.0073	0.067	-0.34	5.6	1.0	-0.4	0.7	0.15	0.000
Fin	1000	0.0071	0.070	0.56	14.3	1.2	5.1	1.3	0.17	0.000
Other	1000	0.0048	0.069	0.36	9.1	1.1	4.3	1.1	0.14	0.000

Table 2: Summary statistics of the monthly excess returns (i.e. $R^k - R^f$)) of the 30 industry portfolios: average; standard deviations; skewness; kurtosis; CAPM-β; γ, the exposure to the skewness risk (see equation 8); δ, the exposure to the kurtosis risk (see equation 9) and the first-order autocorrelation coefficient (with p-value) of portfolio excess returns.

Panel A: Value Decile Portfolios										
Decile	N	Mean	SD	Skew.	Kurt.	β	γ	δ	ACF(1)	p
1=low	1000	0.0054	0.058	-0.02	7.9	1.0	3.0	1.0	0.13	0.000
2	1000	0.0064	0.055	-0.09	8.0	1.0	2.7	0.9	0.09	0.003
3	1000	0.0063	0.054	-0.22	7.8	0.9	2.3	0.9	0.07	0.039
4	1000	0.0062	0.061	1.26	18.8	1.1	6.4	1.3	0.17	0.000
5	1000	0.0069	0.057	0.85	15.3	1.0	5.1	1.1	0.14	0.000
6	1000	0.0073	0.062	0.95	19.2	1.1	5.6	1.3	0.17	0.000
7	1000	0.0074	0.067	1.84	23.4	1.1	8.1	1.4	0.16	0.000
8	1000	0.0090	0.070	2.13	27.2	1.2	8.9	1.5	0.19	0.000
9	1000	0.0098	0.076	1.33	17.3	1.2	7.5	1.5	0.14	0.000
10=high	1000	0.0106	0.094	2.41	27.3	1.5	11.1	1.9	0.16	0.000

Panel B: Size Decile Portfolios										
Decile	N	Mean	SD	Skew.	Kurt.	β	γ	δ	ACF(1)	p
1=small	1000	0.0115	0.103	3.71	39.6	1.4	11.9	1.8	0.22	0.000
2	1000	0.0096	0.090	2.27	25.0	1.4	8.5	1.6	0.19	0.000
3	1000	0.0095	0.082	1.94	23.3	1.3	8.1	1.5	0.22	0.000
4	1000	0.0090	0.076	1.56	18.8	1.3	7.3	1.4	0.19	0.000
5	1000	0.0086	0.073	1.16	16.1	1.2	6.6	1.4	0.18	0.000
6	1000	0.0085	0.070	1.04	15.1	1.2	6.5	1.4	0.18	0.000
7	1000	0.0081	0.066	0.81	14.0	1.2	5.7	1.3	0.16	0.000
8	1000	0.0074	0.062	0.76	13.8	1.1	5.4	1.2	0.14	0.000
9	1000	0.0069	0.059	0.57	13.4	1.1	4.8	1.2	0.12	0.000
10=big	1000	0.0056	0.051	0.09	9.4	0.9	3.1	0.9	0.09	0.006

Panel C: Momentum Decile Portfolios										
Decile	N	Mean	SD	Skew.	Kurt.	β	γ	δ	ACF(1)	p
1=low	995	0.0001	0.099	1.84	19.2	1.6	11.1	1.9	0.16	0.000
2	995	0.0040	0.083	1.84	23.2	1.3	8.9	1.7	0.15	0.000
3	995	0.0041	0.071	1.53	21.8	1.2	7.4	1.5	0.13	0.000
4	995	0.0055	0.065	1.55	20.5	1.1	7.3	1.4	0.13	0.000
5	995	0.0055	0.061	1.31	20.4	1.0	5.7	1.2	0.11	0.000
6	995	0.0062	0.059	0.76	14.8	1.0	4.9	1.2	0.11	0.001
7	995	0.0070	0.056	0.18	10.4	1.0	3.5	1.0	0.07	0.036
8	995	0.0082	0.054	0.04	7.7	0.9	2.6	0.9	0.09	0.006
9	995	0.0089	0.057	-0.30	6.6	1.0	1.5	0.8	0.06	0.061
10=high	995	0.0121	0.066	-0.51	5.2	1.0	0.1	0.7	0.08	0.016

Table 3: Summary statistics of the monthly excess returns (i.e. $R^k - R^f$) of the size, value, and momentum decile portfolios: average; standard deviations; skewness; kurtosis; CAPM-β; γ, the exposure to the skewness risk (see equation 8); δ, the exposure to the kurtosis risk (see equation 9), and the first-order autocorrelation coefficient (with p-value) of portfolio excess returns.

into 17 (30) industry sectors, based on the SIC codes of the previous year.[6] The advantage of industry portfolios is that while similar companies are in the same industry category, the differences between the different industries are considerable. Industry portfolios, therefore, give a broad overview of the economy. An alternative way to group firms into different portfolios is to take some criterion and then form decile portfolios. Doing so results in a large spread of the chosen criterion between the portfolios. Black et al. (1972) were the first to use this method by grouping portfolios based on the past CAPM-$\beta = \operatorname{cov}(r_m, R^e)/\operatorname{var}(r_m)$ of the assets. Later, Fama and French (1992) and many others used value and size portfolios to study size and value anomalies. However, this method may also yield spurious results because of data snooping (see Lo and MacKinlay (1990) and Conrad et al. (2003)). From the Fama-French data library, the value, size, and momentum decile portfolios are used. The 10-value portfolios are formed every July by means of sorting the book-to-market ratio of the previous year. The size decile portfolios for July until the following June are based on the market capitalization in June of the previous year from all available assets listed on the NYSE, AMEX, and NASDAQ. The momentum decile portfolios are calculated based on the returns between $t-2$ and $t-12$. To retain an asset in a momentum portfolio, the prices in $t-13$ and the capitalization of that asset in $t-1$ must be available. An extension of Black et al. (1972) is to use higher moment risk factors instead of the CAPM risk factor, β, for forming decile portfolios. The aim of this procedure is to obtain assets that have risk exposures that are as different from higher order risk as possible. Analogous to Kraus and Litzenberger (1976) and Post et al. (2008), decile portfolios based on the skewness risk

$$\gamma^i = \frac{\mathbb{E}\left[\left(R^M - \mathbb{E}(R^M)\right)^2 \left(R^i - \mathbb{E}(R^i)\right)\right]}{\mathbb{E}\left[\left(R^M - \mathbb{E}(R^M)\right)^3\right]} \tag{8}$$

and additionally to them on the kurtosis risk

$$\delta^i = \frac{\mathbb{E}\left[\left(R^M - \mathbb{E}(R^M)\right)^3 \left(R^i - \mathbb{E}(R^i)\right)\right]}{\mathbb{E}\left[\left(R^M - \mathbb{E}(R^M)\right)^4\right]} \tag{9}$$

are formed, where γ_i and δ_i are generalizations of the CAPM-β for the higher moments risk factors. The γ and δ portfolios are formed every July based on

[6] The data and a detailed description of the industry sectors can be found in the Fama-French data library at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

the data of the previous 36 months. The deciles are calculated on the NYSE data. The portfolios contain all the common stocks from the NYSE, AMEX, and NASDAQ, where the returns of the previous 36 months are available.

An overview of the return characteristics can be found in Tables 1 to 4. The first column includes the number of observations for each portfolio. For every portfolio, all available data points are used, so the number of observations varies slightly. Except for the momentum portfolios, all Fama-French data range from July 1926 to December 2009. The momentum portfolios begin in January 1927, as the returns for the preceding 12 months are needed to calculate momentum. Thereturn data from γ and δ portfolios are available from July 1929 to December 2009. This shorter time horizon stems from the 36-month formation period for those portfolios.

The average returns for small companies and for stocks with a low book-to-market ratio and a high past performance are better, as Figure 1 depicts. These effects were to be expected and were documented in such previous studies as Stattman (1980), Banz (1981), Fama and French (1992, 1993), and Jegadeesh and Titman (1993). A lower coskewness and a higher co-kurtosis also result in higher returns. The return variations between high and low γ and δ risk are considerably smaller than the value, size, and momentum effects. This may indicate that γ and δ are poor indicators for the future skewness and kurtosis risks of the assets.

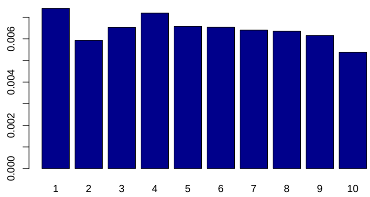
For the 30 industry portfolios, the correlation between β and γ is 0.797, the correlation between β and δ is 0.899, and the correlation between γ and δ is 0.941. This means that a large part of the information on the higher moments is already in the lower moments and may indicate that incorporating higher order moments may not add much additional information. This high correlation is also the reason why no grouping with β-portfolios is included in the analysis: it adds no additional information.

The last two columns of the tables provide the first order autocorrelation of monthly returns and the p-value for the null hypothesis that there is no autocorrelation. Significant autocorrelation can be found in the returns in a large majority of the assets. Therefore, Newey-West autocorrelation corrected standard errors will be used for all test statistics in the empirical analysis.

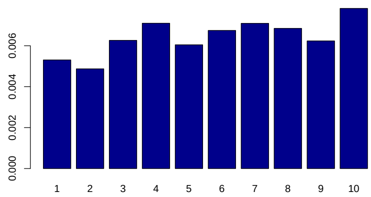
Panel A: γ-Decile Portfolios										
Decile	N	Mean	SD	Skew.	Kurt.	β	γ	δ	ACF(1)	p
1=low	964	0.0074	0.060	-0.20	6.4	0.9	0.8	0.8	0.06	0.070
2	964	0.0059	0.056	-0.40	8.0	0.9	1.5	0.8	0.12	0.000
3	964	0.0065	0.055	0.25	9.2	0.9	3.2	0.9	0.07	0.034
4	964	0.0072	0.057	0.76	12.1	1.0	4.5	1.1	0.10	0.002
5	964	0.0066	0.057	0.63	11.7	1.0	4.2	1.0	0.12	0.000
6	964	0.0065	0.059	0.14	7.8	1.0	3.0	1.0	0.07	0.034
7	964	0.0064	0.062	0.51	12.7	1.1	4.1	1.2	0.11	0.000
8	964	0.0064	0.069	0.80	14.8	1.2	5.4	1.3	0.14	0.000
9	964	0.0062	0.078	1.54	20.7	1.3	7.9	1.6	0.16	0.000
10=high	964	0.0054	0.087	0.99	15.4	1.5	6.7	1.6	0.14	0.000

Panel B: δ-Decile Portfolios										
Decile	N	Mean	SD	Skew.	Kurt.	β	γ	δ	ACF(1)	p
1=low	964	0.0053	0.043	-0.37	7.8	0.7	1.0	0.6	0.14	0.000
2	964	0.0049	0.047	-0.22	10.6	0.8	1.8	0.8	0.14	0.000
3	964	0.0063	0.050	0.22	10.0	0.8	2.9	0.9	0.07	0.022
4	964	0.0071	0.057	0.95	14.4	1.0	5.0	1.1	0.09	0.004
5	964	0.0060	0.060	0.23	9.1	1.0	3.2	1.0	0.08	0.015
6	964	0.0067	0.067	0.44	11.1	1.2	4.2	1.2	0.08	0.011
7	964	0.0071	0.071	0.70	12.0	1.2	5.4	1.3	0.07	0.026
8	964	0.0068	0.076	0.56	11.8	1.3	5.2	1.4	0.13	0.000
9	964	0.0062	0.087	0.97	14.4	1.5	7.1	1.6	0.12	0.000
10=high	964	0.0078	0.100	1.00	12.6	1.7	8.3	1.8	0.13	0.000

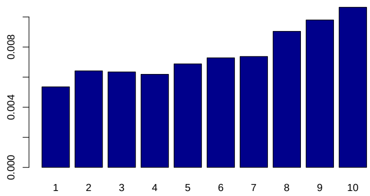
Table 4: Summary statistics of the monthly excess returns (i.e. $R^k - R^f$) of the γ and δ decile portfolios: Average; standard deviations; skewness; kurtosis; CAPM-β; γ, the exposure to the skewness risk (see equation 8); δ, the exposure to the kurtosis risk (see equation 9) and the first-order autocorrelation coefficient (with p-value) of portfolio excess returns.



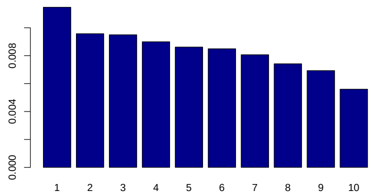
(a) Coskewness: γ



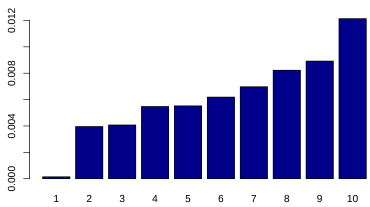
(b) Cokurtosis: δ



(c) Value



(d) Size



(e) Momentum

Figure 1: Average monthly excess returns (i.e. $R^k - R^f$) of the different decile portfolios

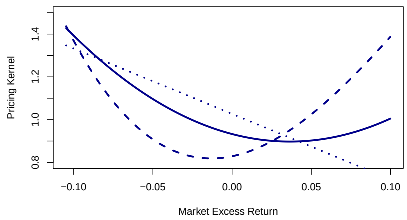
5 Empirical analysis

In this section, the empirical results will be discussed. In a first step, polynomial kernels up to order three are estimated. Quadratic and cubic kernels turn out to have increasing parts if estimated on industrial portfolio data. To check for the significance of the increasing parts of the kernel, the next step is to remove the increasing parts by means of a flat line. However, this makes the fit to the data poorer. A further possibility to estimate kernels is to increase the order of the polynomial further. However, except for the return data of the momentum portfolio, there is no evidence for a kernel of higher order. For the momentum portfolio data, the kernel then turns out to be clearly U-shaped. The estimation of a piecewise linear kernel and further robustness checks will confirm the previous results.

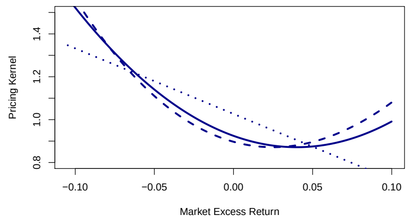
5.1 Linear, quadratic and cubic pricing kernels

Figure 2 provides all of the estimated pricing kernels. The quadratic and cubic kernels are estimated by GMM. For purposes of an independent comparison, the linear benchmark CAPM kernel is also shown. The cubic and quadratic pricing kernels are similar, indicating that the cubic kernel does not behave in a totally different way from the quadratic kernel. The quadratic pricing kernels are all positive, and the shape is generally convex (for the cubic kernel evidence is more mixed), which is in line with the representative expected utility maximizer with decreasing marginal utility. Not in line with that are the increasing parts of some pricing kernels. The range of the x-axis of these figures is chosen carefully: the range between -0.105 and 0.1, which covers 95% of the market returns observed. Outside of this range, the number of observations is small; therefore, the estimation of the kernel is imprecise. A broader range on the x-axis would make the U-shapes obviously more impressive; nonetheless, only a few observations would exist in that additional area, and the kernel estimates would not be very reliable. The increasing regions observed in the industrial portfolios are because of areas where enough observations for reliable estimation are available. For the other portfolios, no evidence for increasing parts is available.

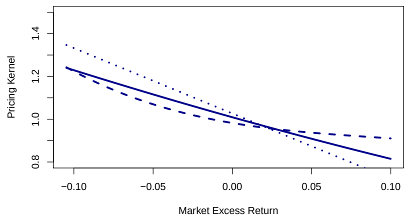
Polynomial pricing kernels up to order 3 are estimated in Table 5 by GMM. The J-statistic shows that a linear pricing kernel (the CAPM) is misspecified—that is, the moment conditions are statistically different from zero. Only for the dataset with the value and size portfolios does a linear pricing kernel appear to be appropriate. For the linear model with the γ



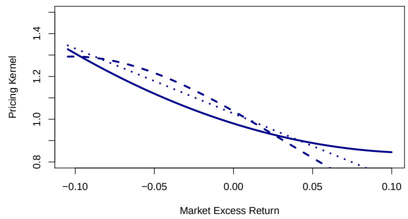
(a) 17 industries



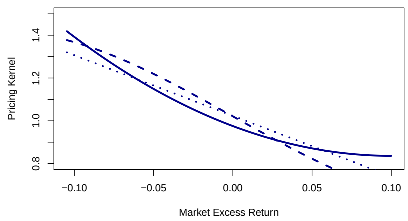
(b) 30 industries



(c) Value and size



(d) Value, size and momentum



(e) γ and δ

Figure 2: Estimated polynomial pricing kernels. The full line is the quadratic pricing kernel; the dashed line is the cubic pricing kernel, and the dotted line is the benchmark CAPM kernel. Estimation details for the quadratic and cubic kernel can be found in subsection 3.3 and in subsection 3.1 for the CAPM benchmark.

and δ portfolios, which should especially take into account the higher-order risk, the J-statistic is only weakly significant at the 10% level in the linear specification. This may indicate two things: either γ and δ are poor indicators for the higher-order risk of the next year, or there is not much nonlinearity in the pricing kernel.

The J-statistic for the quadratic kernel of the momentum portfolio is still significant at the 1% level. Up to this point, everything else appears to be reasonably specified with a quadratic kernel. For example, as Kraus and Litzenberger (1976) also show, the quadratic term is positive for all portfolios. However, the quadratic parameters are not very significant. These t-statistics are in line with Potì and Wang (2010), who find the polynomial terms of order 2 or more to be insignificant. The Wald test, which tests the null hypothesis that the model is linear, shows that for the industry portfolios, a linear kernel can be rejected at the 10% significance level. The linear term in the quadratic kernel is negative, as expected from the CAPM. Moving to a cubic kernel reveals no improvement in terms of the J-statistic, and the sign of the cubic parameter is ambiguous. The Wald test also does not show large differences from a linear model. The likelihood ratio test in Table 10 shows that for the 17 industry portfolios, the cubic model is almost significantly different from the quadratic model at the 1% level. Moreover, the parameter for the cubic term is different from zero at the 10% level. For the other portfolios, there appears to be no reason to move to a cubic kernel.

Up to the momentum portfolio with its highly significant J-statistic, all models can be reasonably well estimated by a polynomial up to order 3. Some increasing regions are found in the kernel of the industry portfolios. The next step is to examine the increasing regions in more detail.

5.2 A closer look at the increasing regions of the kernel

If the kernel is not just linear, a quadratic pricing kernel has an increasing region. Therefore, it is first checked whether or not that region is in an area that includes some observations of r_m. If there are no observations in that area, the increasing region is irrelevant; and if there are only a few market returns in the increasing region, then the result is most likely a statistic artifact. R_{min}, the minimum of a quadratic pricing kernel, can be determined by setting the first derivative of the kernel to zero:

$$R_{min} = \frac{-\theta_1}{2\theta_2}. \tag{10}$$

Panel A: Linear Kernel (CAPM)

	Ind17		Ind30		VS		VSM		γ and δ	
θ_0	1.01	**	1.02	**	1.01	**	1.01	**	1.01	**
	(0.01)		(0.01)		(0.01)		(0.01)		(0.01)	
θ_1	-2.17	**	-2.33	**	-2.20	**	-1.61	**	-2.01	**
	(0.62)		(0.60)		(0.59)		(0.57)		(0.59)	
J	33	**	51	**	20		67	**	27	

Panel B: Quadratic Kernel

	Ind17		Ind30		VS		VSM		γ and δ	
θ_0	0.93	**	0.92	**	1.01	**	0.98	**	0.98	**
	(0.04)		(0.04)		(0.02)		(0.02)		(0.03)	
θ_1	-1.94	*	-2.65	**	-2.07	**	-2.31	**	-2.78	**
	(0.99)		(0.96)		(0.66)		(0.73)		(0.86)	
θ_2	26.65	†	33.26	*	1.27		9.66		13.87	
	(15.05)		(15.40)		(5.12)		(6.91)		(10.13)	
J	23		40	†	20		55	**	18	
W	3.1	†	4.7	*	0.06		2.0		1.9	

Panel C: Cubic Kernel

	Ind17		Ind30		VS		VSM		γ and δ	
θ_0	0.83	**	0.90	**	0.98	**	1.04	**	1.02	**
	(0.08)		(0.07)		(0.06)		(0.04)		(0.05)	
θ_1	1.53		-2.05		-1.25		-4.18	**	-4.13	*
	(1.89)		(1.54)		(1.68)		(1.23)		(1.62)	
θ_2	54.99	*	42.53	†	8.48		-8.05		0.02	
	(27.85)		(23.95)		(18.19)		(12.20)		(17.06)	
θ_3	-143.31	†	-36.46		-32.18		82.43	†	66.85	
	(79.89)		(69.43)		(63.28)		(43.18)		(59.60)	
J	17		40	†	20		64	**	17	
W	3.9		5	†	0.26		4.4		2.7	

Table 5: GMM Estimation of the polynomial pricing kernels, $L = \theta_0 + \theta_1 r_m + \theta_2 r_m^2 + \theta_3 r_m^3$, for the 17 and 30 industry portfolios (Ind17 and Ind30), for the 10-value and 10-size portfolios (VS), for the 10-value, 10-size, and 10-momentum portfolios (VSM) and the γ and δ portfolios. Estimation details can be found in subsection 3.3. The Wald statistic tests if the quadratic (cubic) kernel is different from the linear kernel. †, *, and ** indicate significance at the 10%, 5% and 1% levels, respectively.

Kernel	Variable	Ind17	Ind30	VS	VSM	γ and δ
Quadratic	R_{min}	0.036	0.040	0.818	0.119	0.100
	$p(r_m \geq R_{min})$	0.279	0.250	0.000	0.016	0.025
Cubic	R_{min}	0.036	0.040	0.818	0.119	0.100
	$p(r_m \geq R_{min})$	0.279	0.250	0.000	0.016	0.025
	R_{max}	0.036	0.040	0.818	0.119	0.100
	$p(r_m \geq R_{max})$	0.279	0.250	0.000	0.016	0.025

Table 6: Global minimum (maximum), that is, the turning points of the quadratic and cubic pricing kernels and fraction of the values of the market return that are larger than the turning points.

In the cubic case, the function can have up to one local minimum and one local maximum, which are given by:

$$R_{extrema} = \frac{\theta_2 \pm \sqrt{\theta_2^2 - 3\theta_1\theta_3}}{3\theta_3}.$$

Table 6 provides the minima of the quadratic kernel and the local extrema of the cubic kernel. The plausibility of increasing parts of the kernel is measured by the probability that a market return is in the increasing area; for the quadratic kernel this is, for example, $p(r_m \geq R_{min})$. This probability is measured by the number of months the kernel was in an increasing area divided by the number of all observations. For the industry portfolios, a monthly return larger than 3.6 and 4%, respectively, is sufficient for belonging to the increasing part of the kernel. This implies that in more than 25% of all time periods, the realized pricing kernel was in the increasing region. This observation is supported by the results from the cubic kernel. Not much evidence of an increasing kernel can be found in the other portfolios. With, at most, 2.6% of all months, the quadratic pricing kernel was increasing. For the value/size portfolios, the cubic kernel implies that there are no local extrema—that is, the pricing kernel is decreasing everywhere. For the value, size and momentum, and γ and δ portfolios, the local maxima are in extremely negative returns, and the local minima in extremely positive returns. Therefore, the kernel is falling.

A next step is to check if these increasing parts are statistically significant. One way to do that is to test if the minimum of the quadratic kernel is within the observed data (or at least in an area where almost no data are observed). If this can be rejected, the pricing kernel has increasing parts in the relevant range of market portfolio returns. The null hypothesis is, therefore, that the minimum of a quadratic pricing kernel is at r_{min}. Equation (10) implies for

r_{min}	$P(r_m \geq r_{min})$	Ind17	Ind30	VS	VSM	γ and δ
0.1	0.0250	0.2859	0.2139	0.0597	0.7984	0.9977
0.2	0.0050	0.1536	0.0870	0.4042	0.5713	0.4783
0.3	0.0030	0.1224	0.0622	0.6470	0.3947	0.3476
0.4	0.0000	0.1090	0.0523	0.7845	0.3212	0.2932

Table 7: P-values for a Wald test of the null hypothesis that r_{min} is the global minimum of the quadratic pricing kernel. The second column shows the empirical likelihood that the market return is larger than or equal to r_{min}. The test is given for the 17 and 30 industries; the value and size; the value, size, and momentum; and the γ and δ portfolios.

the null hypothesis that

$$\theta_1 + 2r_{min}\theta_2 = 0.$$

The null hypothesis of $R_{min} = r_{min}$ is tested for r_{min} of 10, 20, 30, and 40%. In the case of $r_{min} = 0.1$, only 2.5% of all market excess returns are larger than r_{min}, and in the case of $r_{min} = 0.4$, no observed market excess return is larger. If the minimum of the pricing kernel is at one of these levels, increasing parts of the pricing kernel are in areas with (almost) no observations. Therefore, they would be irrelevant; that is, if the null hypothesis of the test cannot be rejected, increasing parts in the kernel cannot be significantly statistically supported. Table 7 includes the p-values of the Wald test. As shown, in no case is the estimated minimum of the quadratic pricing kernel different from r_{min} at the 5% level. A model with an increasing kernel is, therefore, not significantly different from one without: that is, the increasing parts are not significant.

A quadratic kernel always has an increasing part. To ensure that this part is not just an artifact from the functional form, a new kernel is used. The basic kernel has a quadratic form, but the slope of the kernel right to the minima is set to zero: that is, after the minimum, the kernel becomes a flat line, as illustrated with the dashed line in Figure 3 for the estimation for the 30-industry dataset. The estimated parameters of these kernels are in Panel A of Table 8. The parameters themselves are similar to the quadratic kernels in Table 5. However, the value of the J-statistic in four of the five portfolios is larger than in the quadratic case. The J-statistic is the sum of the weighted quadratic moment deviation divided by the number of time periods—that is, the criterion minimized by GMM. The smaller values indicate that the increasing parts in the pricing kernel improve the fit of the model.

The earlier approach can be generalized. A maximal level of the slope of

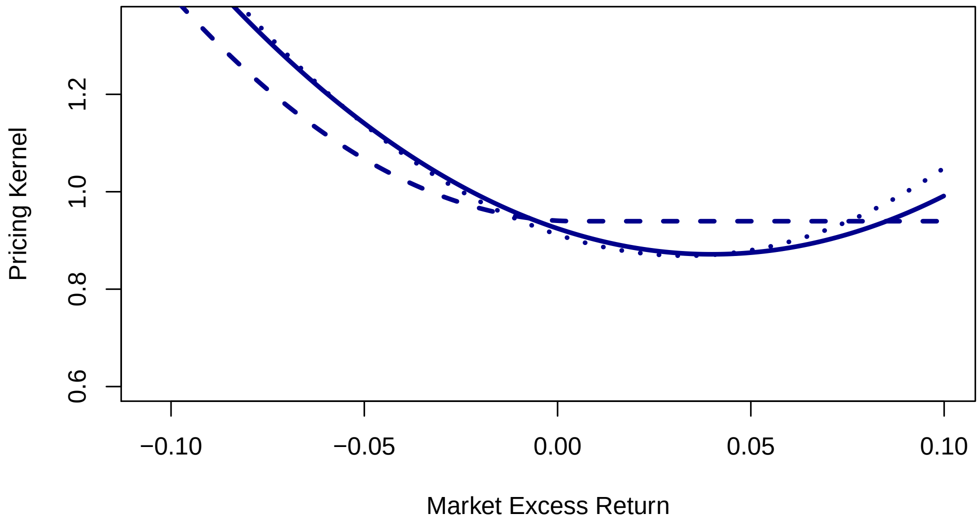


Figure 3: Estimated pricing kernels for the 30 industry portfolios. The full line is the quadratic pricing kernel; the dashed line is the quadratic kernel, which is flat after the minimum; and the dotted line is the quadratic kernel, which continues linear after the slope 8.71. Estimation details can be found in subsection 3.3.

the kernel, m, is fixed. If the slope of the estimated quadratic kernel would be larger than m, the slope is set to m. With $m = 0$, the kernel is as previously, and if $m = +\infty$, the kernel is a standard quadratic kernel. If the kernel were decreasing, as the representative agent model suggests, one would expect the kernel to continue at one point in the decreasing part of the U with a linear negative slope: that is, $m \leq 0$. $m > 0$ (given that $\theta_2 > 0$) indicates the opposite: the linear part starts after the minima and is increasing. The dotted line in Figure 3 illustrates this. With $m = 8.71$, the positive slope starts far outside the plotted range; therefore, the kernel is U-shaped. An estimate of m can be found in the lower part of Table 8. m is positive in four of the five portfolios such that the kernel contains an increasing part. Nonetheless, m is not significantly different from zero; that is, there is no statistically significant evidence for increasing parts in the pricing kernel.

Panel A: Quadratic Kernel Flat after Minimum										
	Ind17		Ind30		VS		VSM		γ and δ	
θ_0	2.4		0.94	**	1.01	**	0.96	**	0.99	**
	(5.5)		(0.02)		(0.02)		(0.05)		(0.05)	
θ_1	33.6		-0.42		-2.07	**	-1.51		-2.40	**
	(93.1)		(5.53)		(0.66)		(2.30)		(0.77)	
θ_2	194.9		42.24		1.27		19.86		6.89	
	(337.3)		(50.92)		(5.12)		(31.20)		(19.38)	
J	22		47	*	20		60	**	23	

Panel B: Quadratic Kernel until Slope is m, afterwards Linear										
	Ind17		Ind30		VS		VSM		γ and δ	
θ_0	0.84	**	0.91	**	1.0		1.04	**	0.97	**
	(0.13)		(0.08)		(2.2)		(0.04)		(0.04)	
θ_1	2.61		-2.61	*	-1.5		-4.18	**	-2.91	†
	(8.32)		(1.17)		(42.8)		(1.23)		(1.51)	
θ_2	90.75		38.89		3.6		-8.87		15.52	
	(68.65)		(29.54)		(154.2)		(12.11)		(13.95)	
m	3.80		8.71		-1.9		82.16	†	7.55	
	(4.37)		(7.90)		(1.6)		(42.95)		(68.62)	
J	20		40	†	20		63	**	18	

Table 8: Panel A shows the estimation of a quadratic pricing kernel when the kernel becomes flat and when the quadratic function is minimal. Panel B shows a quadratic kernel in which the slope is restricted to be smaller or equal to m. If the slope is in a certain range larger than m, the kernel is made linear with a slope of m in that range. Estimations were done for the 17 and 30 industry portfolios (Ind17 and Ind30), for the 10-value and 10-size portfolios (VS); for the 10-value, 10-size, and 10-momentum portfolios (VSM); and for the γ and δ portfolios. Estimation details can be found in subsection 3.3. The Wald statistic tests if the quadratic (cubic) kernel is different from the linear kernel. †, *, and ** indicate significance at 10%, 5%, and 1% levels, respectively.

Order	Ind17	Ind30	VS	VSM	γ and δ
1	0.010	0.009	0.466	0.000	0.143
2	0.114	0.079	0.389	0.002	0.543
3	0.342	0.065	0.358	0.000	0.490
4	0.264	0.014	0.412	0.000	0.373
5	0.249	0.028	0.364	0.000	0.405
6	0.233	0.018	0.262	0.256	0.299
7	0.469	0.044	0.237	0.275	0.678

Table 9: P-values of J-test: The null hypothesis H_0 is that the polynomial pricing kernels of orders 1 to 7 are able to explain the moment conditions.

5.3 Higher-order pricing kernel

Up to now, only polynomials up to the third order have been taken into account. Higher-order polynomials may reveal even more information. In the next step, polynomials up to order 7 are considered. Table 9 gives the p-values of the J-statistics that check if the moment conditions are satisfied for the different polynomials. For quadratic and cubic kernels, the J-statistic is not significant for most portfolios. Only for the value, size, and momentum portfolio does a kernel above order 3 help. In that instance, the J-statistics are highly significant until order 5 and not significant after orders 6 and 7. In the case of the 30 industry, the J-statistics turn out to be insignificant for a quadratic kernel but significant for some higher-order polynomials. Intuitively, one would expect that a higher-order kernel would always fit the data better than a lower order kernel and, therefore, that the J-statistic would fall with the order of the polynomial since a higher-order kernel is, by definition, always able to fit the data at least as well as a kernel of higher order. However, the more parameters the model has, the less over-identifying restrictions exist; therefore, the degrees of freedom of the χ^2 distribution of the J-statistic become smaller with the higher polynomial order of the kernel, and this has a decreasing effect on the p-value. The first effect is typically stronger, and, therefore, the p-values are rising most of the time with the order of the polynomial of the kernel.

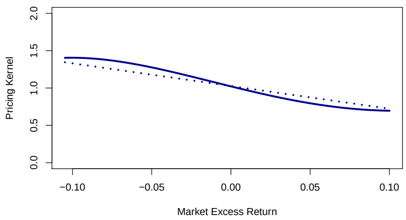
If a model with a kernel of a higher order does not perform better than one with a lower order, there is no reason to choose the model with more parameters. The likelihood ratio test checks if two nested models are statistically different. If the higher-order model is not different, then it is better to choose the lower-order model. Table 10 shows that the linear kernel can be rejected in almost all cases against the higher-order kernels. The exception is the value size portfolios. Linear kernels are, therefore, not sufficient. Except

Panel A: H_0 is a Linear Kernel					
Order	Ind17	Ind30	VS	VSM	γ and δ
2	0.001	0.001	1.000	0.001	0.003
3	0.000	0.004	0.859	0.229	0.010
4	0.001	0.123	0.528	0.067	0.037
5	0.002	0.044	0.641	0.193	0.040
6	0.003	0.100	0.870	0.000	0.095
7	0.001	0.026	0.868	0.000	0.016

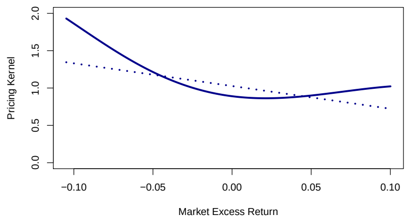
Panel B: H_0 is a Quadratic Kernel					
Order	Ind17	Ind30	VS	VSM	γ and δ
3	0.012	0.648	0.470	1.000	0.646
4	0.047	1.000	0.295	1.000	1.000
5	0.073	1.000	0.433	1.000	0.804
6	0.098	1.000	0.724	0.000	0.987
7	0.031	0.648	0.743	0.000	0.253

Panel C: H_0 is a Cubic Kernel					
Order	Ind17	Ind30	VS	VSM	γ and δ
4	1.000	1.000	0.166	0.040	1.000
5	0.733	1.000	0.330	0.208	0.677
6	0.689	1.000	0.673	0.000	0.987
7	0.204	0.537	0.700	0.000	0.173

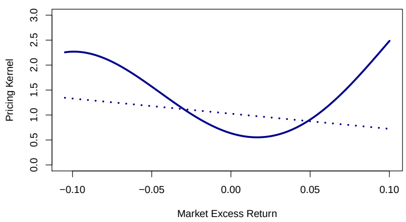
Table 10: P-values of likelihood ratio tests. The linear, quadratic, and cubic kernels are tested against pricing kernels up to order 7. The null hypothesis H_0 is the linear, quadratic, or cubic pricing kernel.



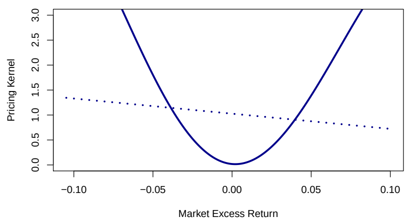
(a) Order 4



(b) Order 5



(c) Order 6



(d) Order 7

Figure 4: Pricing kernels for the value, size, and momentum portfolios for polynomial kernels of orders 4 to 7. The dotted line is the benchmark CAPM. Estimation details can be found in subsection 3.3.

for the data with the momentum portfolios, all other datasets can be modeled with a cubic kernel in the case of the 17-industry dataset and a quadratic kernel for the other portfolios. In the case of the momentum portfolios, the kernels with orders 6 and 7 are always different from the kernels with orders 1 to 3. The kernels of orders 4, 5, 6, and 7 as estimated with the momentum data are shown in Figure 4. The kernels of orders 4 and 5 have a significant J-statistic (i.e., are misspecified) and are not significantly different from the kernels of orders 2 and 3. Put differently, they have practically no increasing parts (for that see figure 2(d)). The two kernels with the higher order are statistically significantly different from the lower-order kernels and are well specified. The momentum portfolio, therefore, requires a kernel of at least order 6. The momentum kernel is U-shaped and contains an increasing part. The lower-order kernels, which are unable to explain the average returns of the momentum portfolios, did not contain an increasing part, showing also that a U-shaped kernel is needed to explain the risk premium on the momentum portfolios.

In line with Dittmar (2002) and Potì (2006), a kernel up to order 3 is required in most cases, considering the J-statistic and the likelihood ratio test. The only exception is the dataset with the momentum portfolio. However, the fact that the U-shape of the kernel becomes massively stronger with the use of a higher order kernel even strengthens the hypothesis of U-shaped kernels.

5.4 Piecewise linear kernel

Up to this point, the focus has been on polynomials. A major problem, especially with the quadratic kernel, is that there are increasing parts of the kernel almost by construction. To verify that these increasing parts are not an artifact of the chosen functional form, piecewise linear kernels are estimated:

$$L = \tau_0 + \tau_1 r_m + \begin{cases} 0 & \text{for } r_m < q_1 \\ \tau_2(r_m - q_1) & \text{for } q_1 \leq r_m < q_2 \\ \tau_2(r_m - q_1) + \tau_3(r_m - q_2) & \text{for } q_2 \leq r_m < q_3. \end{cases} \tag{11}$$

In this setup, there must not be any increasing part in the kernel. In the following, the market portfolio returns are split into three quantiles (with 33% of the observations of r_M in each), and the kernel is estimated. The estimations can be found in Table 11, and the piecewise linear kernel is plotted together with the quadratic kernel in Figure 5. The specification tests are quite similar to those for the quadratic and cubic kernels. The only

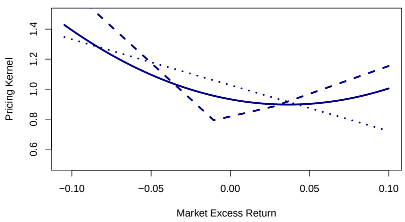
	Ind17		Ind30		VS		VSM		γ and δ	
τ_0	0.69	*	1.00	**	0.91	**	1.01	**	1.60	**
	(0.32)		(0.27)		(0.33)		(0.25)		(0.26)	
τ_1	-9.67		-5.39		-3.75		-3.80		6.39	
	(6.37)		(5.15)		(5.92)		(4.59)		(4.33)	
τ_2	12.11		-4.04		5.49		-2.93		-30.30	*
	(15.87)		(13.32)		(16.25)		(12.04)		(12.16)	
τ_3	1.28		15.52		-4.53		9.62		26.33	**
	(12.05)		(11.13)		(11.31)		(8.87)		(10.05)	
J	22		41	†	21		54	**	14	
W	2.9		4.7	†	0.2		3.5		7	*

Table 11: Estimation of the piecewise linear pricing kernel (i.e., equation (11)), for the 17 and 30 industry portfolios (Ind17 and Ind30), the 10-value and 10-size portfolios (VS), the 10-value, 10-size, and 10-momentum portfolios (VSM), and the γ and δ portfolios. Estimation details can be found in subsection 3.3. The Wald statistic tests whether or not the quadratic (cubic) kernel is different from the linear kernel. †, *, and ** indicate significance at 10%, 5%, and 1% levels, respectively.

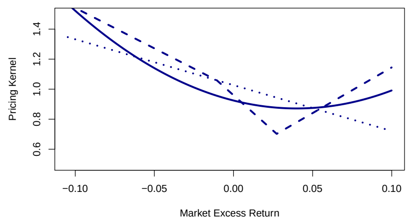
exception is the kernel estimated from the γ and δ portfolios, which start with an increasing part, fall, and then increase again. This change is strong enough that a Wald test indicates that the model is statistically different from a linear model. In the plots, all kernels are more or less moving around the quadratic kernel. Four of the five portfolios show an increasing kernel in the third quantile. If this slope is significantly positive, it can be checked using a Wald test. The p-values for the 17 and 30 industry, the VSM, and the γ and δ portfolios are 0.367, 0.142, 0.327, and 0.459. This shows again that the kernel might be U-shaped, and also that on this occasion, statistical significance is an issue.

5.5 Robustness checks

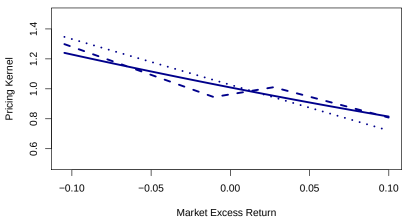
All the estimations have been made using five sets of portfolios, which can be seen as a first robustness check. A next obvious robustness check is to pool all datasets (i.e., 30 industries, value, size, momentum, and γ and δ-portfolios); the results were comparable to the results of the 30 industries-portfolio. Using more assets also did not improve the significance of the results. Further, using the identity matrix, the covariance matrix of the returns or a standard covariance matrix (not taking into account the serial correlation) as the inverse of the weighting matrix $\boldsymbol{W}$ does not change the



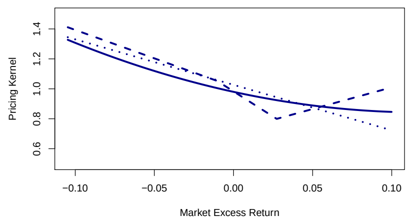
(a) 17 industries



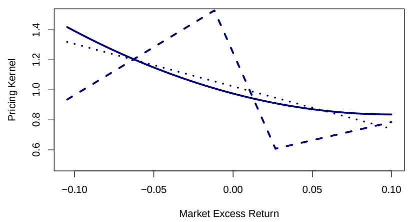
(b) 30 industries



(c) Value and size



(d) Value, size and momentum



(e) γ and δ

Figure 5: Piecewise liner pricing kernel (dashed line) compared with the quadratic pricing kernel (full line) and the benchmark CAPM (dotted line). Estimation details can be found in subsection 3.3.

Time period	Ind17	Ind30	VS	VSM	γ and δ
$\leq$ 1950		0.232			
$p(r_m \geq R_{min})$		0.014			
1951-1970	0.020	0.478		0.028	
$p(r_m \geq R_{min})$	0.417	0.000		0.346	
1971-1990		0.113	0.047	0.015	0.053
$p(r_m \geq R_{min})$		0.025	0.221	0.454	0.154
1991-2009	0.084	0.019	0.012	-0.010	
$p(r_m \geq R_{min})$	0.017	0.435	0.527	0.674	

Table 12: Global minimum of the quadratic pricing kernel and fraction of the values of the market return larger than the turning point. If the estimated quadratic term had a negative sign, the cell has been left empty.

general shape of the pricing kernels. The estimations via GMM and OLS are, furthermore, similar. The next step is to check the time stability and to check if there are any issues with multicollinearity.

The increasing parts of the pricing kernels are the main points of interest. A good indicator for these is the global minima of a quadratic pricing kernel. In Table 12, these are calculated for several time windows of approximately 20 years. In cases where the quadratic function had a maximum instead of a minimum, the cells are left open. For the period before 1950, including the Great Depression and World War II, all kernels are almost linear (exact coefficients are not tabulated), and most of the second-order coefficients are slightly negative. In the case of the 30 industry portfolios, the quadratic term is slightly positive with a value of 3.64. This implies a turning point at 23.2% market returns per month. The probability that this or an even larger return occurs is 1.4%, which is extremely small. In the other much less extraordinary time periods, most pricing kernels are convex and, therefore, have a minimum. In over one-half of the sub periods after 1950, the probability of being in an increasing part of the kernel exceeds 20%. In contrast to the estimation over the whole time period in Table 5, the momentum portfolio already shows a quadratic kernel with increasing parts. For the whole period, a kernel of at least order 6 was needed to see this. However, the evidence for increasing parts in the pricing kernel becomes much smaller for the industry portfolios. Overall, there is evidence for increasing pricing kernels after 1950.

Another potential issue is multicollinearity, since r_m, r_m^2, and r_m^3 are by definition correlated. To address this issue, estimations with orthogonalized

	Ind17		Ind30		VS		VSM		γ and δ	
Panel A: Quadratic Kernel (Orthogonalized)										
θ_0	0.93	**	0.92	**	1.01	**	0.98	**	0.98	**
	(0.04)		(0.04)		(0.02)		(0.02)		(0.03)	
$\bar{\theta}_1$	-1.94	*	-2.66	**	-2.07	**	-2.31	**	-2.78	**
	(0.99)		(0.96)		(0.66)		(0.73)		(0.86)	
$\bar{\theta}_2$	26.71	†	33.33	*	1.32		9.72		13.94	
	(15.06)		(15.40)		(5.13)		(6.91)		(10.14)	
J	23		40	†	20		55	**	18	
W	3.1	†	4.7	*	0.07		2.0		1.9	

	Ind17		Ind30		VS		VSM		γ and δ	
Panel B: Cubic Kernel (Orthogonalized)										
θ_0	0.83	**	0.90	**	0.98	**	1.03	**	1.02	**
	(0.08)		(0.07)		(0.06)		(0.04)		(0.05)	
$\bar{\theta}_1$	1.40		-2.11		-1.22		-4.14	**	-4.16	*
	(1.89)		(1.53)		(1.69)		(1.24)		(1.62)	
$\bar{\theta}_2$	55.57	*	41.88	†	8.82		-5.77		-0.49	
	(27.97)		(23.72)		(18.23)		(12.41)		(16.97)	
$\bar{\theta}_3$	-150.48	†	-39.74		-34.83		81.66	†	68.55	
	(84.22)		(71.80)		(66.29)		(44.86)		(61.45)	
J	17		40	†	20		65	**	18	
W	4.0		5	†	0.28		4.4		2.7	

Table 13: Estimation of the orthogonalized pricing kernel, $L = \theta_0 + \bar{\theta}_1 r_m + \bar{\theta}_2 \left(r_m^2 - a r_m\right) + \bar{\theta}_3 \left(r_m^3 - b r_m^2 - c r_m\right)$, kernels for the 17 and 30 industry portfolios (Ind17 and Ind30); the 10-value and the 10-size portfolios (VS); the 10-value, 10-size, and-10 momentum portfolios (VSM); and the γ and δ portfolios. Estimation details can be found in subsection 3.3. The Wald statistic tests whether or not the quadratic (cubic) kernel is different from the linear kernel. Estimation details can be found in subsection 3.3. †, *, and ** indicate significance at 10%, 5%, and 1% levels, respectively.

regressors are performed. That is,

$$L = \theta_0 + \bar{\theta}_1 r_m + \bar{\theta}_2 \left(r_m^2 - a r_m\right) + \bar{\theta}_3 \left(r_m^3 - b r_m^2 - c r_m\right).$$

is the estimated kernel. a, b and c are defined such that

$$\begin{aligned} 0 &= \operatorname{cov}(r_m, r_m^2 - a r_m) = \operatorname{cov}(r_m, r_m^3 - b r_m^2 - c r_m) \\ &= \operatorname{cov}(r_m^2, r_m^3 - b r_m^2 - c r_m). \end{aligned}$$

The results for the estimation of the pricing kernel with these orthogonalized factors (or polynomials) can be found in Table 13. As shown, these are comparable with the results in Table 5–that is, they show that the pricing kernel for the nonmomentum portfolios must be around orders 2 or 3. In addition, the signs of the polynomials are identical. The results are, therefore, robust for multicollinearity.

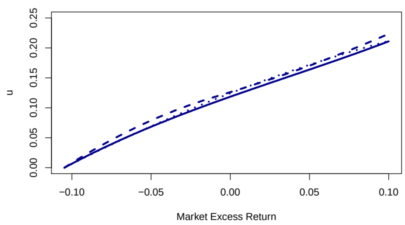
6 Utility function of the representative agent

Assuming there is a representative agent, what would his utility function look like? In equation (4), it was shown that the pricing kernel is a constant times the marginal utility, i.e. $L(r_m) = \delta \cdot u'(r_m)/u'(c_0)$. Integrating this implies

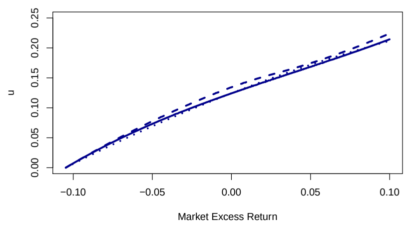
$$u(r_m) = \text{constant} + \frac{\delta}{u'(c_0)} \int_{-\infty}^{r_M} L(r) dr.$$

This analysis requires that markets be complete and that the representative agent has correct beliefs and an increasing, concave utility function. None of these assumptions must be satisfied. Nevertheless, it is interesting to see the shape of the utility function that would evolve from these assumptions.

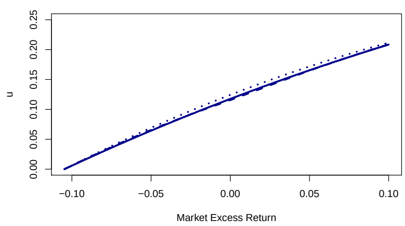
If one normalizes $\delta/u'(c_0) = 1$ and sets the utility function at the left corner of the graph to zero, the utility function implied for the quadratic, the piecewise linear, and the benchmark CAPM kernel can be found, as shown in Figure 6. The most obvious point is that all utility functions are quite similar. This fits the fact that it is difficult to find statistically significant differences between the different kernels. Nonconcavities are especially observed with the piecewise linear pricing kernel. The problem is that the shape of the utility function, as implied by the piecewise linear kernel, is different for every portfolio. The quadratic pricing kernel is typically much closer to the CAPM benchmark. Nonconcavities are observed in the case of the industry portfolios, but even there they seem to be weak. Overall, the effect of the nonconcavities in the utility function seems to be weak. Their existence could, nonetheless, alter completely the investment behavior of the representative agent.



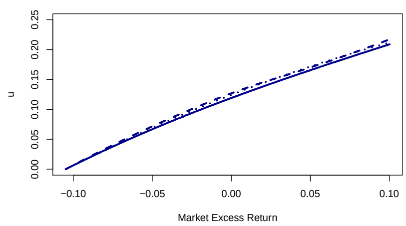
(a) 17 industries



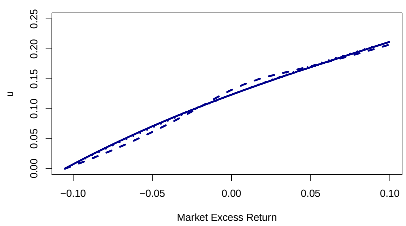
(b) 30 industries



(c) Value and size



(d) Value, size and momentum



(e) γ and δ

Figure 6: Utility function of the representative agent implied by the pricing kernel in the case of the quadratic pricing kernel (full line), the piecewise linear pricing kernel (dashed line), and the benchmark CAPM (dotted line).

7 Conclusions

This paper examined the increasing parts of the U-shaped pricing kernels found in equity data. This has been done in a much broader way than can be found in the existing literature. In particular, the estimation on datasets in the industry and momentum portfolios shows clear evidence for increasing parts in the pricing kernel. To make sure that these increasing parts are not just an artifact of the polynomial functional form, other functional forms that allow for nonincreasing shapes lead to a poorer fit for the data. Despite the fact that the U-shape of the kernel can be shown on many different datasets, time horizons, and functional forms in terms of statistical significance, this evidence is weak. This paper shows that analogously to factor models, the value, size, and momentum effect can be explained by the polynomials of market returns of sufficiently high order. Another contribution is that the kernels of these higher-order polynomials are mainly U-shaped, increasing with positive returns. This is consistent with a positive premium on coskewness.

An implication of the increasing part of the pricing kernel is that the economy cannot be modeled by a risk-averse, utility-maximizing representative agent. This paper shows that this effect is not just a short-run phenomenon, as with the evidence from stock options data that typically holds for a specific, typically short, time period. The increasing parts in the kernel appear to persist over a time horizon of more than 80 years. So far, there is no generally accepted economic explanation for this phenomenon. For future research, it will be important to check to which degree heterogeneous, misestimated beliefs, Peso problems, incomplete markets, aggregation problems, and nonstandard preferences contribute to the empirically observed U-shape. To date, there is clear evidence that it is not possible to explain the whole phenomenon with only one of those factors.

References

Stanislav Anatolyev. GMM, GEL, Serial Correlation, and Asymptotic Bias. *Econometrica*, 73(3):983–1002, May 2005.

Yacine Aït-Sahalia and Andrew W. Lo. Nonparametric Risk Management and Implied Risk Aversion. *Journal of Econometrics*, 94(1-2):9–51, January–February 2000.

Rolf W. Banz. The Relationship between Return and Market Value of Common Stocks. *Journal of Financial Economics*, 9(1):3–18, 1981.

Nicholas Barberis and Ming Huang. The Loss Aversion / Narrow Framing Approach to the Equity Premium Puzzle. NBER Working Papers 12378, National Bureau of Economic Research, Inc, July 2006.

Giovanni Barone-Adesi. Arbitrage Equilibrium with Skewed Asset Returns. *Journal of Financial and Quantitative Analysis*, 20(3):299–313, September 1985.

Giovanni Barone-Adesi and Hakim Dall'O. Is the Price Kernel Monotone? Working paper, University of Lugano, January 2010.

Giovanni Barone-Adesi, Robert F. Engle, and Loriano Mancini. A GARCH Option Pricing Model with Filtered Historical Simulation. *Review of Financial Studies*, 21(3):1223–1258, May 2008.

Fischer Black, Michael C. Jensen, and Myron S. Scholes. The Capital Asset Pricing Model: Some Empirical Tests. In Michael C. Jensen, editor, *Studies in the Theory of Capital Markets*. Praeger Publishers Inc, 1972.

Douglas T. Breeden, Michael R. Gibbons, and Robert H. Litzenberger. Empirical Test of the Consumption-Oriented CAPM. *Journal of Finance*, 44 (2):231–262, June 1989.

Michael J. Brennan and Yihong Xia. Dynamic Asset Allocation under Inflation. *The Journal of Finance*, 57(3):1201–1238, June 2002.

David A. Chapman. Approximating the Asset Pricing Kernel. *Journal of Finance*, 52(4):1383–1410, September 1997.

Pierre Chaussé. Computing generalized method of moments and generalized empirical likelihood with R. *Journal of Statistical Software*, 34(11):1–35, May 2010.

Y. Peter Chung, Herb Johnson, and Michael J. Schill. Asset Pricing When Returns Are Nonnormal: Fama French Factors versus Higher Order Systematic Comoments. *Journal of Business*, 79(2):923–940, March 2006.

John Howland Cochrane. *Asset Pricing*. Princeton University Press, 2001.

Jennifer Conrad, Michael Cooper, and Gautam Kaul. Value versus Glamour. *Journal of Finance*, 58(5):1969–1995, October 2003.

Enrico De Giorgi and Thierry Post. Second-Order Stochastic Dominance, Reward-Risk Portfolio Selection, and the CAPM. *Journal of Financial and Quantitative Analysis*, 43(02):525–546, June 2008.

Kai Detlefsen, Wolfgang Härdle, and Rouslan Moro. Empirical Pricing Kernels and Investor Preferences. SFB 649 Discussion Papers 2007-017, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany, April 2007.

Robert F. Dittmar. Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns. *Journal of Finance*, 57(1):369–403, February 2002.

Friedhelm Eicker. Limit Theorems for Regressions with Unequal and Dependent Errors. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 59–62. University of California Press, 1967.

Vihang R. Errunza and Oumar Sy. A Three-Moment International Asset-Pricing Model: Theory and Evidence. Working paper, McGill University, November 2005.

Eugene F. Fama and Kenneth R. French. The Cross-Section of Expected Stock Returns. *Journal of Finance*, 47(2):427–465, June 1992.

Eugene F. Fama and Kenneth R. French. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33(1): 3–56, February 1993.

Hsing Fang and Tsong-Yue Lai. Co-Kurtosis and Capital Asset Pricing. *Financial Review*, 32(2):293–307, May 1997.

Wayne E. Ferson and Campbell R. Harvey. Seasonality and Consumption-Based Asset Pricing. *Journal of Finance*, 47(2):511–552, June 1992.

Irwin Friend and Randolph Westerfield. Co-Skewness and Capital Asset Pricing. *Journal of Finance*, 35(4):897–913, September 1980.

Yuri Golubev, Wolfgang Härdle, and Roman Timofeev. Testing Monotonicity of Pricing Kernels. SFB 649 Discussion Papers 2008-001, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany, January 2008.

James D. Hamilton. *Time Series Analysis*. Princeton University Press, 1. edition, 1994.

Lars Peter Hansen. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029–54, July 1982.

Lars Peter Hansen and Kenneth J. Singleton. Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models. *Econometrica*, 50(5):1269–1286, September 1982.

Lars Peter Hansen and Kenneth J. Singleton. Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns. *Journal of Political Economy*, 91(2):249–265, April 1983.

Lars Peter Hansen, John Heaton, and Amir Yaron. Finite-Sample Properties of Some Alternative GMM Estimators. *Journal of Business & Economic Statistics*, 14(3):262–280, July 1996.

Michael J. Harrison and David M. Kreps. Martingales and Arbitrage in Multiperiod Securities Markets. *Journal of Economic Theory*, 20(3):381–408, June 1979.

Campbell R. Harvey and Akhtar Siddique. Conditional Skewness in Asset Pricing Tests. *Journal of Finance*, 55(3):1263–1295, June 2000.

Thorsten Hens and Christian Reichlin. Three Solutions to the Pricing Kernel Puzzle. Swiss Finance Institute Research Paper 10-14, Swiss Finance Institute, March 2010.

Thorsten Hens and Peter Wöhrmann. Mental Accounting and the Equity Premium Puzzle. Working Paper 243, NCCR "Financial Valuation and Risk Management", July 2006.

Chi-Hsiou Hung. Return Predictability of Higher-Moment CAPM Market Models. *Journal of Business Finance & Accounting*, 35(7-8):998–1022, September 2008.

Jens Carsten Jackwerth. Recovering Risk Aversion from Option Prices and Realized Returns. *Review of Financial Studies*, 13(2):433–451, Summer 2000.

Narasimhan Jegadeesh and Sheridan Titman. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, 48(1):65–91, March 1993.

Alan Kraus and Robert H. Litzenberger. Skewness Preference and the Valuation of Risk Assets. *Journal of Finance*, 31(4):1085–1100, September 1976.

Kian-Guan Lim. A New Test of the Three-Moment Capital Asset Pricing Model. *Journal of Financial and Quantitative Analysis*, 24(2):205–216, June 1989.

John Lintner. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*, 47(1):13–37, February 1965.

Andrew W. Lo and A. Craig MacKinlay. Data-Snooping Biases in Tests of Financial Asset Pricing Models. *Review of Financial Studies*, 3(3):431–467, Autumn 1990.

Michael Magill and Martine Quinzii. *Theory of Incomplete Markets*. MIT Press, 1996.

N. Gregory Mankiw and Matthew D. Shapiro. Risk and Return: Consumption Beta Versus Market Beta. *The Review of Economics and Statistics*, 68(3):452–459, August 1986.

Jan Mossin. Equilibrium in a Capital Asset Market. *Econometrica*, 34(4): 768–783, October 1966.

Whitney K. Newey and Richard J. Smith. Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica*, 72(1): 219–255, January 2004.

Whitney K. Newey and Kenneth D. West. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3):703–708, May 1987.

Duong Nguyen and Tribhuvan N. Puri. Higher-Order Systematic Comoments and Asset Pricing: New Evidence. *Financial Review*, 44(3):345–369, August 2009.

D. F. Nicholls and A. R. Pagan. Heteroscedasticity in Models with Lagged Dependent Variables. *Econometrica*, 51(4):1233–1242, July 1983.

Thierry Post and Pim van Vliet. Downside Risk and Asset Pricing. *Journal of Banking & Finance*, 30(3):823–849, March 2006.

Thierry Post, Pim van Vliet, and Haim Levy. Risk Aversion and Skewness Preference. *Journal of Banking & Finance*, 32(7):1178–1187, July 2008.

Valerio Potì. The Coskewness Puzzle in the Cross-Section of Industry Portfolio Returns. Working paper, Dublin City University Business School, March 2006.

Valerio Potì and DengLi Wang. The Coskewness Puzzle. *Journal of Banking & Finance*, 34(8):1827–1838, August 2010.

Joshua V. Rosenberg and Robert F. Engle. Empirical Pricing Kernels. *Journal of Financial Economics*, 64(3):341–372, June 2002.

Mark E. Rubinstein. The Fundamental Theorem of Parameter-Preference Security Valuation. *Journal of Financial and Quantitative Analysis*, 8(1): 61–69, January 1973.

William F. Sharpe. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, 19(3):425–442, September 1964.

Daniel T. Slesnick. Are Our Data Relevant to the Theory? The Case of Aggregate Consumption. *Journal of Business & Economic Statistics*, 16 (1):52–61, January 1998.

Daniel R. Smith. Conditional Coskewness and Asset Pricing. *Journal of Empirical Finance*, 14(1):91–119, January 2007.

Pascal St-Amour. Benchmarks in Aggregate Household Portfolios. Cahiers de Recherches Economiques du Département d'Econométrie et d'Economie politique (DEEP) 07.07, Université de Lausanne, Faculté des HEC, DEEP, January 2007.

Dennis Stattman. Book Values and Stock Returns. *Chicago MBA: A Journal of Selected Papers*, 4:25–45, 1980.

Halbert White. A Heteroskedasticity-Consistent Covarince Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48(4):817–838, May 1980.

David W. Wilcox. The Construction of U.S. Consumption Data: Some Facts and Their Implications for Empirical Work. *American Economic Review*, 82(4):922–941, September 1992.

Alexandre Ziegler. Why Does Implied Risk Aversion Smile? *Review of Financial Studies*, 20(3):859–904, May 2007.

A Proof of consistency of the pricing kernel estimation via factor model (OLS)

This appendix establishes the consistency of the estimation of the pricing kernel as estimated by the OLS factor model from section 3.2. For consistency, a more specific setup is required:

$$R_t^{e,k} - \mathbb{E}\left(R_t^{e,k}\right) = (\boldsymbol{f_t} - \mathbb{E}(\boldsymbol{f_t}))'\boldsymbol{\beta_k} + \epsilon_t^k \tag{12}$$

$$\mathbb{E}\left(R_t^{e,k}\right) = \boldsymbol{\beta_k'}\boldsymbol{\lambda} + \eta_k^*, \tag{13}$$

where $R_t^{e,k}$ are the excess returns of asset k in period t, $\boldsymbol{f_t}$ is a stochastic vector of factors in period t, and the vector $\boldsymbol{\beta_k}$ is the factor exposures of asset k. The risk premia associated with the factors is the vector $\boldsymbol{\lambda}$. The vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\beta_k}$ for $k = 1, \ldots, K$ are fixed but with unknown parameters to estimate. ϵ_t^k and η_k^* are noise terms. The OLS estimator of $\boldsymbol{\beta_k}$ is $\boldsymbol{\hat{\beta}_k}$ and is estimated using:

$$R_t^{e,k} - \frac{1}{T}\sum_{t=1}^{T} R_t^{e,k} = \left(\boldsymbol{f_t} - \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}\right)'\boldsymbol{\beta_k} + \epsilon_t^k.$$

If the expected value is replaced with the sample average,[7] the second equation can be stated as follows:

$$\frac{1}{T}\sum_{t=1}^{T} R_t^{e,k} = \boldsymbol{\hat{\beta}_k'}\boldsymbol{\lambda} + \eta_k.$$

Using OLS, $\boldsymbol{\lambda}$ can be estimated from this equation by

$$\boldsymbol{\hat{\lambda}} = \left(\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{\hat{\beta}_k}\boldsymbol{\hat{\beta}_k'}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{R_t^e}\right)$$

with $\boldsymbol{R_t^e}$ as the vector of the excess returns of all assets in period t. A consistent estimator of the covariance matrix of $\boldsymbol{f_t}$ is

$$\widehat{\mathrm{Var}}(\boldsymbol{f_t}) = \frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{f_t} - \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}\right)\left(\boldsymbol{f_t} - \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}\right)'.$$

[7]This replacement is unproblematic, as it is possible to include the estimation error only in the error term—i.e., $\eta_k = \eta_k^* - \mathbb{E}\left(R_t^{e,k}\right) + \frac{1}{T}\sum_{t=1}^{T} R_t^{e,k}$.

From Section 3.2, it is known that the parameter of interest is $\boldsymbol{b} = -\operatorname{var}(\boldsymbol{f_t})^{-1}\boldsymbol{\lambda}$. An obvious candidate for an estimator of $\boldsymbol{b}$ is, therefore,

$$\hat{\boldsymbol{b}} = -\widehat{\operatorname{Var}}(\boldsymbol{f_t})^{-1}\hat{\boldsymbol{\lambda}}.$$

The next step is to show that $\hat{\boldsymbol{b}}$ is a consistent estimator. For this, the following assumptions are required:

- $\boldsymbol{f_t}$ is stochastic and $\left(\boldsymbol{f_t} - \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}\right)\epsilon_t^k$ is a martingale difference sequence for all assets k
- $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t} \xrightarrow{p} \mathbb{E}\left(\boldsymbol{f_t}\right)$, where $\left|\mathbb{E}\left(\boldsymbol{f_t}\right)\right| < \infty$
- $\mathbb{E}\left[\left(\epsilon_t^k\right)^2\left(\boldsymbol{f_t} - \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}\right)\left(\boldsymbol{f_t} - \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}\right)'\right] = \boldsymbol{\Sigma_t}$, a positive definite matrix, with $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\Sigma_t}$ converging to a positive definite matrix $\boldsymbol{\Sigma}$ and
$$\frac{1}{T}\sum_{t=1}^{T}\left(\epsilon_t^k\right)^2\left(\boldsymbol{f_t} - \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}\right)\left(\boldsymbol{f_t} - \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}\right)' \xrightarrow{p} \boldsymbol{\Sigma}$$
- $\widehat{\operatorname{Var}}(\boldsymbol{f_t}) \xrightarrow{p} \boldsymbol{Q} = \operatorname{var}(\boldsymbol{f_t})$, where $\boldsymbol{Q}$ is nonsingular
- $\frac{1}{K}\sum_{t=1}^{T}\boldsymbol{\beta_k}\boldsymbol{\beta_k}'$ is a finite nonsingular matrix
- η_k and η_l, $l \neq k$ are independent random variables with $\mathbb{E}(\eta_k) = 0$ and $\operatorname{var}(\eta_k) = \sigma_i^2 < \infty$.
- η_k is furthermore independent of all $\boldsymbol{f_t}$ and ϵ_t^k

The assumptions about $\boldsymbol{f_t}$ and ϵ_t^k are standard assumptions to ensure consistency for the parameters of a regression model with general heteroskedasticity in the error terms; this model dates back to Eicker (1967), White (1980), Hansen (1982) and Nicholls and Pagan (1983). Therefore, the OLS estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}$ for equation (12) converges to $\boldsymbol{\beta_k}$, that is,

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{k}} \xrightarrow{p} \boldsymbol{\beta_k}.$$

If $1/T\sum_{t=1}^{T}R_t^{e,k} = \boldsymbol{\beta_k}'\boldsymbol{\lambda} + \eta_k$ is plugged into $\hat{\boldsymbol{\lambda}}$, the estimator $\boldsymbol{b}$ is

$$\begin{aligned}
\hat{\boldsymbol{b}} = -\widehat{\operatorname{Var}}(\boldsymbol{f_t})^{-1}\hat{\boldsymbol{\lambda}} &= -\widehat{\operatorname{Var}}(\boldsymbol{f_t})^{-1}\left(\frac{1}{K}\sum_{k=1}^{K}\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}'\right)^{-1}\left(\frac{1}{K}\sum_{k=1}^{K}\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}\boldsymbol{\beta_k}'\boldsymbol{\lambda} + \eta_k\right)\\
&= -\widehat{\operatorname{Var}}(\boldsymbol{f_t})^{-1}\left(\frac{1}{K}\sum_{k=1}^{K}\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}'\right)^{-1}\left(\frac{1}{K}\sum_{k=1}^{K}\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}\boldsymbol{\beta_k}'\boldsymbol{\lambda} + \frac{1}{K}\sum_{k=1}^{K}\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}\eta_k\right).
\end{aligned}$$

A first step to prove the consistency of $\hat{\boldsymbol{b}}$ is to show that

$$\frac{1}{K}\sum_{k=1}^{K}\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}\eta_k \xrightarrow{p} 0. \tag{14}$$

$\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}$ is a function of $\boldsymbol{f_t}$ and ϵ_t^k. η_k is independent of these two variables. Therefore, η_k and $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}$ are also independent and $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}\eta_k$ is a martingale difference sequence. The assumptions further imply bounded covariance matrices for η_k and $\boldsymbol{f_t}$. Therefore, $\mathrm{cov}(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}, \eta_k)$ is bounded. Then, by example 7.11 in Hamilton (1994), equation (14) holds. Further, $\widehat{\mathrm{Var}}(\boldsymbol{f_t}) \xrightarrow{p} \mathrm{var}(\boldsymbol{f_t})$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}} \xrightarrow{p} \boldsymbol{\beta_k}$. From the fact that there are only continuous functions in the estimator, it follows that

$$\begin{aligned}
\hat{\boldsymbol{b}} &\xrightarrow{p} -\mathrm{var}(\boldsymbol{f_t})^{-1}\left(\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{\beta_k}\boldsymbol{\beta}_{\boldsymbol{k}}'\right)^{-1}\left(\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{\beta_k}\boldsymbol{\beta}_{\boldsymbol{k}}'\boldsymbol{\lambda}+0\right)\\
\hat{\boldsymbol{b}} &\xrightarrow{p} -\mathrm{var}(\boldsymbol{f_t})^{-1}\boldsymbol{\lambda} = \boldsymbol{b}.
\end{aligned}$$

In other words, $\hat{\boldsymbol{b}}$ is a consistent estimator for $\boldsymbol{b}$. However, the number of assets and the number of time steps have to converge to infinity for this estimator to be consistent.